December 1971

AD

ACN 17494

AD 738180

Development of a Division War Game Model (DIVWAG)

Final Report

ANALYTICAL METHODOLOGIES

Volume II

UNITED STATES ARMY COMBAT DEVELOPMENTS COMMAND

Combat Developments Research Office

Approved for public release - distribution unlimited.

Destroy this report when no longer needed.
Do not return it to the originator.

SEE AD
738179

167

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* COMBAT DEVELOPMENTS RESEARCH OFFICE, COMPUTER SCIENCES CORPORATION	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

Development of a Division War Game Model, Volume II Analytical Methodologies

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Final Report

5. AUTHOR(S) *(First name, middle initial, last name)*

Combat Development Research Office,
Computer Sciences Corporation

6. REPORT DATE December 1971	7a. TOTAL NO. OF PAGES 1[illegible]6	7b. NO. OF REFS 10
8a. CONTRACT OR GRANT NO. DAAG-11-70-C-0875 b. PROJECT NO. USACDC WORK DIRECTIVE 7-71 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) ACN 17494 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*	

10. DISTRIBUTION STATEMENT

Distribution of this document is unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY USACDC COMBAT SYSTEMS GROUP FORT LEAVENWORTH, KANSAS 66027

13. ABSTRACT

This volume describes the development and ~~intended~~ application of the analytical methodologies to be used in conjunction with the war game model. It includes ~~an analytical~~ methodology for analyzing the output of the model(s) developed ~~and~~ determining the effectiveness of a single force, ~~An analytical methodology~~ for comparing alternative forces ~~is also developed~~ in response to statement of work requirements.

DD FORM 1 NOV 65 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

UNCLASSIFIED
Security Classification

14. KEY WORDS	LINK A		LINK B		LIN
	ROLE	WT	ROLE	WT	ROLE
Alternative Forces Analysis of Variance by Ranks Combat Effectiveness Combat Potential Combat Service Support Combat Support Computer Assisted Model Doctrine Evaluation Firepower Force Design Hypothesis Testing Intelligence Measures of Effectiveness Mobility Nonparametric Methodology Sensitivity Testing War Game					

PROJECT NUMBER 7-71

DEVELOPMENT OF A DIVISION WAR GAME MODEL (DIVWAG) VOLUME II-ANALYTICAL METHODOLOGIES

PREPARED FOR

USACDC COMBAT SYSTEMS GROUP

UNDER CONTRACT DAAG-11-70-C-0875

IN RESPONSE TO

USACDC WORK DIRECTIVE 7-71, DEVELOPMENT OF A DIVISION WAR GAME MODEL (DIVWAG), 11 MAY 1971

DECEMBER 1971

CSC

COMPUTER SCIENCES CORPORATION

NOTES

ABSTRACT

This volume is one of a series of documents describing the work performed by the Combat Developments Research Office (CDRO) for the U.S. Army Combat Developments Command (USACDC) in response to USACDC Work Directive 7-71, Development of a Division War Game Model (DIVWAG).

The Combat Developments Research Office is operated by the Field Services Division of Computer Sciences Corporation under Contract DAAG-11-70-C-0875.

NOTES

CONTENTS

CONTENTS (Cont)

CONTENTS (Cont)

Page

Chapter 3 (Cont)

NOTES

FIGURES

FIGURES (Cont)

FIGURES (Cont)

NOTES

CHAPTER 1

FORCE ANALYSIS MANAGEMENT

Section I. INTRODUCTION

1. PURPOSE. The DIVWAG Statement of Work required the design, development, and validation of (1) an analytical methodology for determining the effectiveness of a single force and (2) an analytical methodology for comparing alternative forces. The evaluation and comparison methodologies developed to fulfill these requirements are described in this volume.

2. SCOPE. This chapter provides prospective users of the analytical methodologies with guidance in the management of a task involving the analysis of division forces. The chapter is oriented to the managerial level rather than the functional level to aid the manager of a force analysis task in planning and organizing the task effort to meet the analysis objectives. This chapter emphasizes the use of computer-assisted war gaming in the solution of force analysis problems; but the philosophies and techniques are applicable to the management of analysis tasks employing a variety of analytical methods. Chapter 2 is devoted to the methodology for evaluating the effectiveness of a single force. Chapter 3 describes the methodology for comparing alternative forces. Chapters 2 and 3 are oriented to the functional level managers and provide procedures and techniques for the application of the analytical methods described therein.

3. BACKGROUND:

a. The ability of management to understand the problem to be solved, to design a methodology for its solution, and to develop a plan for the timely and efficient execution of the methodology is fundamental to the success of any force analysis.

(1) The problem to be solved must be analyzed in terms of the objectives or purposes of the entire effort, anticipated use of the study results, and the resources available for application to the study.

(2) Once the study problem is understood and clearly defined, an appropriate and efficient methodology for its solution must be selected or designed. The methodology must be appropriate in terms of providing usable answers or insights to the study problem and efficient in terms of making the best possible use of resources within the constraints of the study.

(3) After the problem and methodology are well defined, the development and application of a management plan become of paramount importance. The plan must provide for efficient use of personnel resources, must consider calendar time constraints, and must ensure that the study effort maintains its direction and that the results are integrated to fulfill study objectives.

b. War games are the primary analytical tools to assist in the orderly examination of conflict situations involving military units and systems. The entire framework of a war game is open for inspection by a force planner so that the applicability of game results to hardware procurement and organizational development can be studied in detail.

c. War games can support research in a variety of ways. They can, even in their formulative stages, provide considerable broad general insight into critical problems in study areas; they can generate distributions of outcomes of play of specific situations; and they can function as pseudo-experiments, producing data for analysis after the plays are completed.

d. The last, analytic, mode is possible only when efforts have been made to ensure that the required elements of game record, or data, are available. Given a basic operational structure of movement, contact, and battle between the opposing resolved units, the approach is one of introducing detailed simulations of the real world events to be studied. These simulations result from cooperative effort between the game staff and the analysts of the proponent agency. Research objectives are used to develop the simulation models, rules of play, and assessment procedures that ensure events pertinent to the problems do occur in the course of play and that the desired data for analysis are taken. One criticism that has been leveled against the conduct of war games is that the analysts have conspicuously failed to reap the rewards in doing analytic research based on data from their games. A point easily lost is that the play of the war game merely produces data for detailed analysis. The data produced from the play of the war game must be interpreted and evaluated to produce insights, findings, and conclusions that are valid for the situation being simulated in the game; reliable in the sense that repeated play of the same plans and set of conditions would yield similar results within the limits of chance variation; and useful for predicting results for related situations. The actual results of the game must be analyzed, and the analysis must also appraise the validity of the input data, the rules and assumptions made, the availability of resources consumed, and the strategy and tactics utilized by the player teams. Such analysis can be a great source of information to the sponsor of the study effort and can help justify the expense of the game. Applying a structural methodology to the output of the war game takes it out of the realm of philosophy and back into the science of operations research.

e. The analytical methodologies described in detail in subsequent chapters are presented in the order in which they will be generally conducted. This order is depicted diagramatically in Figure 1-1. The numbers appearing on the diagram are keyed to the step numbers of the following explanation.

(1) Step 1. The initial step in the application of the methodology will be receipt of inputs for the war game. Principal inputs will be the division force design(s) together with the appropriate scenario, threat, environment, and other information relating to organizations, weapons, and doctrine as required.

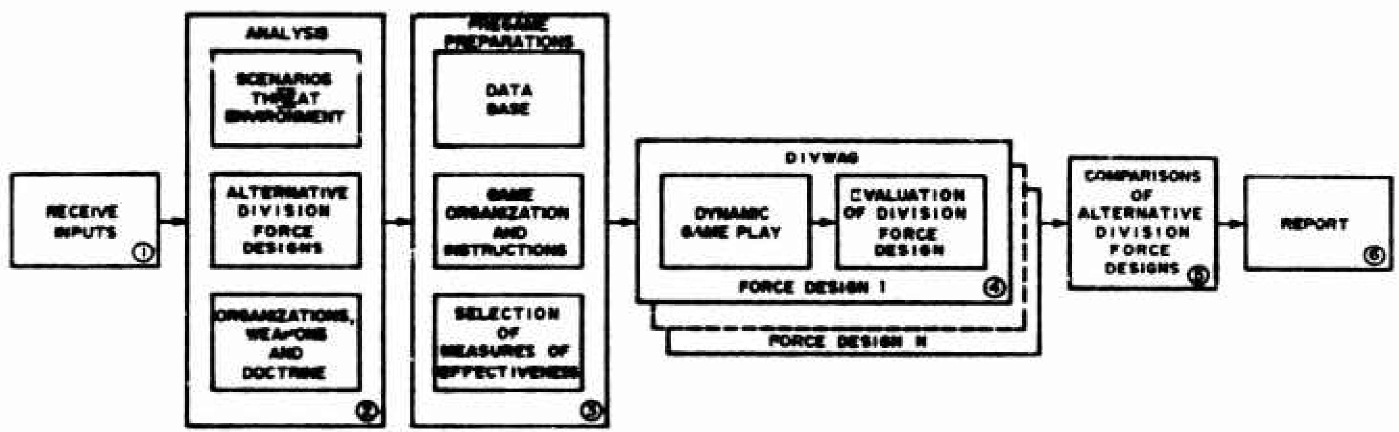


Figure 1-1. Organization of DIVWAG Analytical Methodologies

(2) Step 2. A thorough analysis of the inputs is essential to ensure the elements for analysis are complete, accurate, and understood by the evaluator. The analysis will include, but not be limited to, the enemy situation (threat), geographical area of operations, friendly situation, mission, concept of operations, allied forces situation (if appropriate), constraints, and assumptions.

(3) Step 3. Pregame preparations will ensure compatibility of the objectives and the analytical tools. The data base will be updated and augmented as required for play. The player and controller teams will be organized and oriented. Measures of effectiveness will be selected from the total array of MOE developed to guide the effort.

(4) Step 4. Dynamic game play will commence. Event cycling, as well as periodic evaluation and redirection of game play, will permit more than one game to be played simultaneously, provided game facilities, computer support, and personnel are available. As dynamic play of each alternative force design is completed, the game director, game teams, and operations research and systems analysts will analyze the game results using the force evaluation methodology to ensure the objectives are met and credible output is available. The combat effectiveness of the force will be evaluated using the force evaluation methodology described in Chapter 2.

(5) Step 5. As each division force design is gamed and analyzed, results will be arrayed for comparison with another or other force design(s). Major effectiveness measures of the force designs will be collated and analyzed to establish the strengths and weaknesses of each and to establish relative merit among alternatives. Analyses of the arrayed measures of effectiveness using the comparative methodology will result in a rank ordering of the forces. The comparative methodology is described in detail in Chapter 3.

(6) Step 6. The project team will document the results of pregame analysis and preparation, major facets of game play and force analysis, and results of comparison of alternative division force designs.

4. ORGANIZATION. The preceding paragraphs have provided an overview of the purpose and content of this volume and discussed the requirement for an overall management concept of a force analysis problem. Section II discusses the philosophy and development of war gaming and culminates in a description of the application of computer-assisted war gaming to force analysis problems. Section III discusses the management of a force analysis problem from initial preparation, through production of data, to application of analytical methodologies.

Section II. EVOLUTION OF COMPUTER-ASSISTED WAR GAMING

5. THE PHILOSOPHY OF GAMES:

a. Description of a Game:

(1) A game is a form of human endeavor, sometimes recreational, distinguished from other forms of activity by having rules and a payoff. In return for adhering to the rules, the player receives a reward. The rules are arbitrary, and many payoff schemes exist. Payoff can be determined by chance, as in dice or roulette; a function of skill, as in stud poker, business, or football; or, in the British sense, obtained from playing the game with class (win or lose), as in war. Some games are personally competitive, one player's gain being another player's loss. Competition adds to player interest; however, the most interesting games, with all due respect to the adherents of craps, roulette, and basketball, are the intellectual ones.

(2) Intellectual games have extensions in time, both past and future. Each action by a player produces a new state (situation), and each action is a function of the existing state; therefore, each action by a player is dependent on all the past actions taken by players. In addition, a player is required to project into the future the events that his action will unleash. Reward accrues to the player who can most successfully accomplish the projection. The game becomes an intricate, changing tapestry, which takes on the form of all the past decisions of the players. Chess, war, politics, and billiards are examples of games with extensions in time.

b. Characteristics of a War Game:

(1) A war game is a game having as its goal the replication of one or more of the manifestations of war; that is, the states through which the game passes should be similar, to some degree of detail, to situations encountered in war. The generation of this similarity is accomplished by requiring that the game rules and payoff calculations interact in a way that transitions a starting state, assumed to be realistic, to successive realistic states. Each action by a game player, done in accordance with the game rules, results in a payoff calculation, which in turn produces a new game state. The cycle then repeats (Figure 1-2).

(2) Identification of the rules constitutes a problem in war gaming. If the rules were immutable, they could be written down and machines taught to play a passable game.[1] This technique is, in fact, used for tiny wars;

---

1. Even if the rules can be written down, there is no guarantee that a machine can be taught to play a good game. A prime example is chess. The heuristics of searching the future for a good move are inadequately understood even for this ancient and well-studied game.

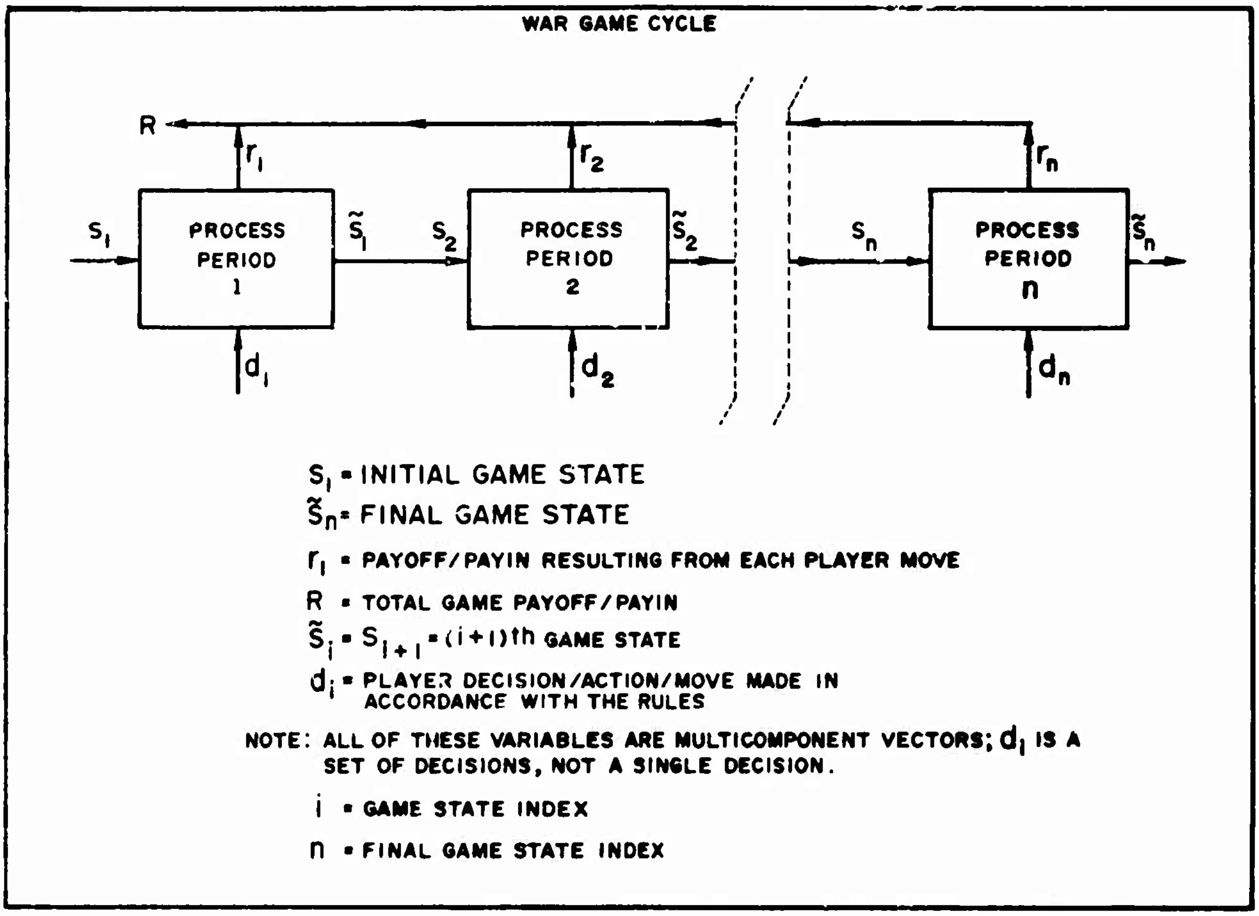


Figure 1-2. War Game Cycle

and simulations, untouched by human hands, come into their own. For grand wars, however, some of the rules for the replicating game are expressed as constraints on player actions. This expression of constraints takes the form of player conformity to what he and the other players regard as generally accepted normal behavior. Normal behavior may admit very few choices for a player decision, or alternatively, a continuous band of choices.

(3) The other rules, those which do not constrain player action, concern translating the existing situation plus the player decision into the next situation. It is important for the player to understand these rules; otherwise, he will have no appreciation of the consequences of his actions, and the final game situation will be nothing more than a random event. It is axiomatic among those who devise the rules for translating the game situation plus decisions into the next situation (i.e., model builders) that players should not know how this is done. It needs to be understood that the requirement for such an axiom implies subterfuge on the part of model builders.

(4) Although machines can be taught to play games, it is generally accepted that use of the word "game" implies human participation. Since some games are used, not for entertainment, but for the study of war, an understanding of the merits of simulations (machine games) and war games (human gamers) becomes necessary. There are four valid reasons for using a human player in the replication of war:

- The player is to be trained.
- A human player is innovative.
- The player rules cannot be formalized to a degree adequate for programming a machine.
- The player rules are so involved that machine programming becomes inefficient and wasteful.

The training of a player is a valid reason only for that class of games designed especially for training purposes. The second reason, the requirement for innovation, has come to be discouraged, the result of a level of sophistication requiring that a series of games be absolutely comparable one with another rather than a sequence in an optimization scheme. Thus, the remaining two reasons for using human players in war gaming are operative; the human player is used from necessity rather than choice. The return for simplifying the model builder's task apparently overshadows the difficulties generated for the person responsible for making comparisons among games.

(5) In practice the necessity for using a human does not exclude the machine player. The process of transitioning a game from one state to the next is gradually being taken over by computers, and it is natural to let the

machine play small games and eliminate some of the arbitrary rules.[2] The result of this trend is affecting the design of war games. More and more the design of a game is becoming a matter of selecting the man-machine interface echelon. Above some echelon, the man makes the decisions; below, the machine makes the decisions.

(6) Despite the ambiguities introduced by a human player, a war game is a valuable analytical procedure. Problems that can be treated in no other way can be studied in a war game. This class includes large problems, complex problems, and problems that are not well understood. The production of a sequence of game states amounts to the fragmenting of a large problem into a set of smaller interconnected problems, and the decisions required at each game state focus the attention of the player on the important elements of the problem. An effective war game produces a number of subproblems, which may be much more amenable to analysis.

(7) A significant characteristic of a war game is that its successful conduct requires a tremendous amount of communication. Often this communication must take place among players with diverse skills and backgrounds. The analyst must explain to the player why he must know certain things, the player must explain military tactics to the rule keeper (model builder), and the whole game must be explained to sponsors who were not players.

(8) War games also have disadvantages, including:

- Results are replicable only with great difficulty.
- Value of a game is a direct function of the skill of the game team. An effective team is difficult to build, and there are no mediocre ones. They are either good or very bad.
- Results are more subject to controversy than those generated by machines. (A computer's subjective judgments are more easily concealed.)
- A game is more subject to external pressure. (Computer decisions can be manipulated also, but only overtly and with difficulty.)

---

2. The elimination of arbitrary rules is good; however, in many cases the rules are based on experience. Replacement of the experience base with a poorly taught machine is not good.

c. Classification of War Games. War games are classified by describing the purpose, the form, and the information constraint.

(1) The purpose of a game may be to train players, to test operational plans, or to research the composition and conduct of forces used in war.[3]

(2) The form of a game may be either free or rigid. These terms are historical and refer to the relative dependence of the game on formalized rules. A rigid game depends entirely on rules. Transition of the game from one state to the other is determined from tables and calculations. The results of player decisions are determined by reference to rules. Opposed to the rigid game is the free game. In the free game a controller decides issues solely on the basis of his judgment. A free game is very fast, more fun, and possibly, nearly as effective as a rigid game.[4] Most war games are a mixture of free and rigid, with a recent steep trend toward the use of rigid games in research.

(3) Classification of a game according to information constraint determines what the players are allowed to know about each other. In a completely open game total information is provided each player. In a closed game only certain elements of the game state of the opposing player are known. Game rules decide what information to provide. It can become a game refinement to let these rules correspond to the information-collecting capability of the player. An open game is advantageous when speed of play or increased management control is required. Although information constraint and speed of play or management control do not appear to be correlated, they are in fact related inexorably for most modern war games. Intelligence play has become an absolute requirement for a war game. Even if a player has 100 percent knowledge of his opponent, an open game, he can act only on the knowledge that he might reasonably be expected to gain in real life. As a result, there is no apparent difference between the events of an open game and the events of a closed game. As a practical matter the difference occurs in the way the gaming staff is organized.

(a) In organizing for an open game, the staff controlling the game and the player staffs are formed into a single committee chaired by the

---

3. A more basic purpose for a war game was expressed by an unnamed WWII General of the Greater German Empire who stated, "The purpose of the exercise was to provide the opportunity for raising and discussing controversial problems with a selected and critical circle." (War Games, Office of Military History, Department of the Army, Washington, D.C., 1952.)

4. J.P. Young, A Survey of Historical Developments in War Games, ORO-SP-98, Operations Research Office, Bethesda, Maryland, 1959.

game director. The players become functional specialists whose purpose is to offer advice. The game decision base becomes flexible and, in the extreme, may consist only of the game director.[5] The result of this organization and the resultant narrowed decision base is a reduction in the requirement for multipoint communication, always a time consumer, and a reduction in the occurrence of the unexpected. The game operates under the guidance of a single intellect. Player sparring and unnecessary activities are eliminated, and the operation is very efficient. An open game, however, has all the characteristics of playing Monopoly with oneself. It is very hard to be creative and very easy to be bored and mediocre. Much can be learned of rules, but very seldom can the intellect be really challenged. An open game is an effective user of time and things, and a very inadequate user of human resources.

(b) The closed game is an almost exact contrapositive of the open game. What the closed game does well, the open game does not, and vice versa. A staff that does well on open games will not do well on closed games. The converse is also true. A closed game is more realistic for the players. They also have their own professionalism at stake. A closed game is generally accompanied by much fire, smoke, and friction, the result of breakdowns in communication. The controlling staff has its work cut out for it, and the game management needs patience. The payoff for this kind of effort is many interesting problems and a great deal of insight. Closed games, therefore, are ideal for optimization schema.

6. DEVELOPMENT OF WAR GAMING:

a. A detailed history of the development of war gaming can be found in an Operations Research Office document entitled A Survey of Historical Developments in War Games.[6] The reader interested in tracing the development and historical applications of war gaming is referred to this excellent source. Basically, the developmental process, which started in prehistory and is not yet finished, has produced three types of war games: battle chess, rigid war games, and free war games.

(1) Battle chess became stylized early in the nineteenth century, and as a result no longer bears much resemblance to the activity it originally represented.

(2) The development of rigid and free type games (defined in Paragraph 5) describes a continuing process of conflict between the requirement for speed and the requirement for realism in gaming war.

---

5. Extreme in that this situation represents one end of a continuous set of possibilities. In the sense of describing the normal situation, it is not extreme. The game usually follows the game director, and the staff implements his decisions.

6. J.P. Young, op. cit.

b. During that period of time when the mathematical arts were making their impact on European culture, the rigid war game was developing. War was believed to be a science that would easily fit into a scheme of rational mechanization. The loss of this original naivete was accompanied by a reaction against the difficult and voluminous rules associated with the rigid war games and the discovery that increases in game complexity achieved only marginal increases in game usefulness. As a result, the free game came into vogue in Europe during the late nineteenth century.

c. In this country a native contentiousness made it difficult to operate a game that did not have formal rules, the difficulty being in finding a controller whose judgments would be recognized as valid by the players. Whether the rules were scrupulously used or not, games with rules that could be invoked as needed tended to become more popular than the free games. The recent conceptualization of the research war game and the advent of a rapid data processing capability have reinforced the ascendancy of the rigid war game.

d. The laws governing the conduct of men and war, and the applicable mathematical operators, are no better understood today than they have been in the past; nevertheless, the art of gaming war has arrived at the computer-assisted war game stage. As in many other fields the ability to store and manipulate information has outpaced the ability to sort the relevant from the irrelevant. The final development of an effective interface between man and machine is still to be achieved.

7. COMPUTER-ASSISTED WAR GAMING:

a. There are two types of computer-assisted war games, which may be referred to as Type A and Type B.

(1) In the Type A game the computer functions only as a very efficient bookkeeper. Its functions are limited to elementary computations and the formatting of feedback data. Decision logic is not employed; thus, the personnel employed in the game retain their decision-making responsibilities.

(2) In a Type B war game the computer takes on all or some of the responsibility for controlling the action of the game and plays some parts of the game. Extensive decision logic is used. Development of Type B games has the goal of restricting human participation to either key issues (those requiring innovation) or decisions absolutely requiring human input (those for which logic is not programmable).

b. A pure Type A or Type B war game does not exist. Existent games are a mixture; however, at the present time most games are predominantly Type A. The rules and calculations of what is philosophically a hand-operated game are programmed, and an interface between players and machine is defined. In most cases rules are extended and more detail considered because of the increased manipulative capability of the computer. This extension of the

rules may take the form of incorporating a simulation into the game. In this case the simulation is merely an elegant formulation of a rule. The game may still be of Type A, because a simulation may or may not describe faithfully a situation requiring a representation of human decision making.

(1) Decision making in a simulation accomplished by using an average decision, a most probable decision, or the selection of a decision in a random fashion from a stated distribution is a disservice to the real thing. When a chain of such decisions is linked together, reality may become the victim.[7]

(2) When a decision is represented by a rule/simulation it is assumed that the stated result always occurs. For many situations this is true; e.g., the squad leader may not have many choices; however, a brigade commander has many options. A simulation of brigade activities, used so that a human player is required to consider only those decisions required at the division echelon, must consider decisions logically.

c. This argument identifies the next step in the development of computer-assisted war games, the realization of the true Type B war game. Human decision making considers future gain. When the future is uncertain, the human player explores the decision tree so that future gain may be estimated. The heuristics of exploring a decision tree must be determined[8] so that the computer can be really taught to game. Computer simulations can then obtain results from decisions based on estimating future gain instead of responding solely to the pressures of the present.

d. The developers and users of computer-assisted war games must exercise considerable caution. The mathematical requirements are formidable. This caution is well stated by Young in his 1959 survey of war gaming:[9]

> Some differences of opinion are being voiced as to whether the rigid types of games, the computer or mathematical models, will ever actually give results which can be applied without a great amount of risk. Many current computer games have been reduced to absurdities because of the simplifying assumptions necessary to develop the mathematics which allow the model to "work." In addition the results of war games are only as good as the input data, and in many cases, because of the continuing development of new weapons and new methods of warfare, such data are lacking or based on frequently optimistic estimates.

Much basic research remains to be done.

---

7. Such functions are termed pathologically bimodal. In translation this term means that the simulation is schizophrenic. The result has an ill-defined relationship to the input.

8. A part of this process is the self-learning program. In such programs the self-learning concerns finding heuristics.

9. J.P. Young, op. cit., p. 103.

Section III. MANAGEMENT OF A DIVISION FORCE ANALYSIS

8. INTRODUCTION. The foregoing discussion of gaming was designed to provide the prospective manager of a division force analysis with insight into the nature of the gaming tools available to him. This section highlights some of the aspects of managing the analysis through three phases:

- Initial preparation for task execution
- Production of evaluation data
- Application of analytical methodologies.

9. INITIAL PREPARATION FOR TASK EXECUTION. The receipt of a work directive from the sponsoring agency initiates preparation for task execution. Management consideration must be directed at several facets, including:

- Analysis of task objectives
- Design of a methodology
- Development of an analysis plan
- Organization of personnel.

a. Analysis of Task Objectives:

(1) The task objectives are normally provided by the sponsoring agency. They embody the purposes for which the study is to be conducted and imply the nature of the expected results. Management must analyze the task objectives to ensure that they are thoroughly understood, unambiguous, and not subject to misinterpretation. The sponsor and the game managers must coordinate very closely during this initial analysis; the sponsor, to ensure that he communicates to the managers the intent and emphasis of the objectives; the managers, to ensure that their interpretation of the objectives coincides precisely with that of the sponsor.

(2) The number and complexity of analysis objectives bear a direct relationship to the study's chances of success. Lt. Gen. Julian J. Ewell (then Major General and Deputy Commanding General, U.S. Army Combat Developments Command) wrote:

> If a major study directive asks ten or fifteen major questions, its chances of a successful ending are heavily compromised before it gets underway. Every effort should be made to narrow a study down to one major question, with four probably the absolute maximum for a reasonable effort and result. The narrowing can

only take place after considerable thought as there are usually many (apparently) logical alternatives or options. However, after much screening effort, the supercilious, redundant, inconsistent, or secondary questions can be determined and either eliminated or placed in a category to be answered only if time permits....Another facet of the same problem is the habit of mixing large and small issues in a directive. This makes a study most difficult. A big issue usually requires a "big grain" study approach, a small issue a "small grain" approach. Mixing them may require two studies in effect or a rather feeble cut at the less important one.[10]

(3) The analysis of task objectives must include an appreciation of the intended use of the analysis results. By remaining aware of the potential applications of study results through all phases of study performance, management can help to ensure that the final product of the analysis meets the sponsor's needs.

(4) The responsibility for the success or failure of a study (and the attendant credit or discredit) rests ultimately with the sponsoring agency; thus, the sponsor is vitally interested in conducting a scientifically and militarily valid study and in obtaining wide acceptance of study results. Task objectives that are too numerous or complex to be addressed within study resources or that do not adequately reflect the intended purpose of the analysis can only lead to unsatisfactory results. Game managers can help to avoid this outcome and to achieve a mutually beneficial end by thoroughly coordinating their analysis of task objectives with the study sponsor and by suggesting redefinition or reorientation of objectives when appropriate; however, the study sponsor cannot abdicate his responsibility to pose the study problem within realistic parameters and to establish task objectives that fairly define that problem.

b. Design of a Methodology. The selection of a methodology by which the objectives will be fulfilled is one of the most crucial elements in the management of the analysis. The methodology must be appropriate to the problem, able to be performed within task resources, and capable of fulfilling task objectives. Chapter 2 of this volume describes in detail the steps involved in developing a methodology for a division force analysis. The steps include, in addition to analysis of objectives:

---

10. Letter, CDCDG to Chief of Staff, USACDC, dated 12 February 1968, subject: Informal Thoughts on Study Management at the USACDC Level.

- Selection of measures of effectiveness and effectiveness indicators
- Definition of performance data requirements and loading of game input data
- Development of a game plan.

These steps in the development of a methodology are described from the management viewpoint in the following subparagraphs.

(1) Selection of Measures of Effectiveness and Effectiveness Indicators:

(a) The primary measures of effectiveness (MOEs), or the criteria upon which the forces will be evaluated, should be apparent from the analysis of task objectives. A faulty selection of the primary MOE reflects an incomplete understanding of the objectives; it can seriously degrade study acceptability since the primary MOE provides the basis for the entire evaluation. On the other hand, the secondary MOEs and effectiveness indicators supporting the primary MOE must be chosen from a wide range of possibilities; their selection entails value judgments as well as a careful analysis of the components of the primary MOE.

(b) Chapter 2 of this volume contains a detailed discussion of the selection of an MOE hierarchy to support a division force analysis. The primary MOE is designated as mission accomplishment, and secondary MOEs are designated for each of the functional areas of land combat. Effectiveness indicators supporting each secondary MOE are chosen on the basis of quantifiable data considered pertinent to the analysis of force effectiveness in the functional area represented by the secondary MOE.

(c) Management must emphasize the importance of the MOE hierarchy as the basis for the analysis methodology. Its selection requires careful and thorough study by analysts familiar with the forces to be evaluated, the doctrine and tactics to be employed, and the evaluation objectives. The MOEs and effectiveness indicators selected must be coordinated with the sponsor and his review board to ensure that they adequately reflect the desired study emphasis.

(2) Definition of Performance Data Requirements and Loading of Game Input Data. A major effort in the preparation for task execution is the definition of performance data requirements and the preparation of a data base to exercise the models selected to produce performance data.

(a) The definition of performance data requirements is an outgrowth of the selection of MOEs and effectiveness indicators. The data necessary to quantify the MOEs and effectiveness indicators for all units

or systems of interest under a prescribed range or combination of conditions constitute the performance data required for the analysis. Management depends upon the study sponsor to provide the parameters for the evaluation through a scenario and other guidance (see subparagraph (3) below); staff analysts then determine performance data needed to conduct the complete evaluation based on the MOEs and effectiveness indicators.

(b) At this point a vehicle for generating the required data can be selected. War games, field tests, simulations, or a wide variety of other techniques might be considered. Some of the advantages and disadvantages of the use of computer-assisted war gaming to produce performance data were discussed in Section II to this chapter. For clarity and simplicity of presentation, the discussion from this point assumes the use of computer-assisted war gaming techniques; the principles are applicable to the management of tasks utilizing a number of other techniques.

(c) The collection of a data base and the loading of input data are critical and time-consuming steps in task preparation. The accumulation of a data base involves:

- Identifying input data requirements for all submodels
- Identifying and accumulating source documents
- Obtaining sponsor approval of data sources
- Verifying that the data are appropriate for their intended use in the model
- Documenting the source and application of all data input
- Preparing input forms, coding, and loading data into the computer.

Management should be aware that each of these steps is time-consuming and that difficulties can arise, especially when data must be obtained for new or conceptual units or systems. A potential for human error exists at several points in the data preparation process, and management should establish a system of checks and approvals to minimize the chances of a damaging data input error. Acceptance of the final analysis product can be jeopardized by unacceptable input data; for this reason management cannot overemphasize the importance of the data collection process and its thorough coordination with the sponsoring agency.

(3) Development of a Game Plan. The conduct of a computer-assisted war game must be preceded by a detailed analysis of the factors that are critical to game operation and the subsequent development of a game plan.

(a) The critical factors analysis can begin upon receipt from the sponsoring agency of detailed data on the organization and equipment of opposing forces, environmental data, and a scenario, which includes the opening situation, guidance, assumptions, and constraints. Management then identifies calendar time constraints imposed upon project performance, obtains an estimate of computer time allocation, and notes other resource constraints, such as manpower, facilities, or equipment.

(b) Within this framework of resource constraints, management then identifies and conducts an analysis of other critical factors. Critical factors will vary from game to game, but they typically include:

- Game requirements
- Game content
- Analysis requirements
- Model operation
- Time constraints
- Game operation.

1. Game Requirements. Game requirements include the number of games to be played, the number of game days to be played, and the specific type engagements required.

2. Game Content. Game content refers to such elements as the forces to be gamed, the level of resolution and degree of aggregation, organizational and equipment considerations, and battle termination criteria.

3. Analysis Requirements. This factor refers to the critical variables to be considered and the quantitative and qualitative data analysis requirements of the evaluation.

4. Model Operation. This factor includes the characteristics of the models to be used, consideration of use of component submodels in a simulation mode, and man/machine interfaces.

5. Time Constraints. The analysis of critical time constraints will consider the number of units to be gamed, the number of game days to be played, calendar time allocated, computer running time to game time ratio, game period turnaround time, manning levels and skills available, and analysis requirements.

6. Game Operation. This subject is resolved after consideration of the impacts of all other critical factors and results generally in the cycle scheduling required for the timely completion of dynamic game, analysis, and documentation effort.

(c) After completion of the critical factors analysis, management can prepare a game plan to guide the staff effort through the conduct of the game. The game plan consists of three basic sections--Game Setting, Rules and Procedures, and Staff Organization--which are discussed in the following subparagraphs.

1. Game Setting. This section of the game plan provides the staff with all essential information concerning the game objectives, the forces to be gamed, and the beginning political/military situation. It may include administrative information, such as security procedures, a glossary of terms, and a reference list.

2. Rules and Procedures. Rules may be of two kinds, game rules and technical rules. Game rules are rules for the conduct of military operations during the game or for making decisions concerning such operations (doctrine and tactics). Technical rules refer to model-dependent considerations. Procedures, on the other hand, are administrative rules developed for the efficient conduct of the game.

3. Staff Organization. This section of the game plan describes the functions and responsibilities of each element of the staff. It should be sufficiently flexible to allow added detailed descriptions of the tasks of individual members of each group.

(d) Figure 1-3 is an outline for a typical game plan. It shows major paragraph headings and gives a brief description of the type information to be included under each heading.

c. Development of an Analysis Plan:

(1) A sound plan for analyzing the performance data produced by the war game model is essential to the successful completion of the study. A haphazard, poorly-planned, or hurried analysis can, at worst, discredit a study or, at best, give rise to charges that the analysis has failed to fully exploit the potential of the war game output.

(2) A plan for analysis of output data should provide for the performance of both subjective and statistical analysis and the integration of results into a product responsive to task objectives. Chapters 2 and 3 of this volume describe a methodology for analyzing force performance data to evaluate a single force and to compare alternative forces. This methodology employs subjective judgment and statistical techniques to derive and test inferences regarding force performance under a variety of background conditions.

(3) Management should ensure that an analysis plan is developed and approved well in advance of the initiation of game play. Then, analysis of data can be performed according to the plan concurrently with play of the game. In this way, analysts can monitor the adequacy of game output for analysis purposes and can identify requirements for side analysis.

Section I. GAME SETTING

1. PURPOSE OF THE PLAN

A statement that the game plan is the basic document providing policy, procedural, organizational, and administrative guidance for the conduct of all game phases.

2. STATEMENT OF THE PROBLEM

A concise statement of the job to be performed and the methodology to be employed.

3. GAME OBJECTIVES

A statement of the specific game objectives as interpreted from the game directive issued by the sponsoring activity.

4. SCENARIO

A description of the conflict situation to be gamed.

5. GENERAL SITUATION

A description of the geographical and political environment for the conflict to be gamed.

6. BLUE (RED) SPECIAL SITUATION

A special situation description for each force to be gamed (Red and Blue) and the disposition and missions of each. Contains privileged information and is issued only to the appropriate player team and to control.

7. GLOSSARY OF TERMS

A list of terms and their definitions applicable to the game plan.

8. REFERENCES

A listing of documents to be used in data base preparation and as doctrinal guidance for the Red and Blue forces to be evaluated.

9. SECURITY PROCEDURES

Standard operating procedures for handling classified defense information, privileged information within the game, and visitors and physical security at the war game facility.

Figure 1-3. Sample Game Plan Outline (continued next page)

Section II. GAME RULES AND PROCEDURES

10. RULES

a. Doctrine

Describes rules on doctrine. Particularly pertinent when new and untested concepts, force structures, and equipments are gamed.

b. Tactics

Describes tactical rules. Impacts particularly on comparability in a project requiring comparison of forces in different games.

c. Technical

Rules reflecting model-dependent considerations.

d. Rules for Decision Making

Applicable specifically to the control group in a rigid game.

11. PROCEDURES

a. Game Cycle

Procedures to be applied in conducting an entire game cycle (as distinguished from a computer cycle). Defines game cycle in terms of beginning and ending point and explains schedule for completion of a normal cycle. Particularly important in keeping project on programmed calendar schedule.

b. Gaming Rate

Provides schedule of game cycles in terms of physical effort of game turnaround and average game time per cycle.

c. Levels of Resolution/Aggregation

Prescribes levels of resolution and aggregation for mainstream game. Also for side analyses and sensitivity tests, if requirements have been identified.

d. Side Analyses/Sensitivity Tests

Procedures for side analyses and sensitivity tests, if such requirements are identified pregame.

Figure 1-3. (continued)

e. Critical (Significant) Events

Any military phenomenon of the battlefield the occurrence of which results in a decided advantage for one of the protagonists. Explains requirements to game such events arising from game directive or from deduction resulting from preliminary force analysis.

f. Personnel Schedules and Roles

Describes any peculiar scheduling or personnel requirements, such as an irregular work schedule for a particular period because of computer hour allocation.

g. Computer Interface

Establishes procedures for assembly of game cycle turnaround data into machine readable format and prescribes gaming element responsible for assembly, delivery to and pickup from the computer, and the associated records of the entire process.

h. Records Requirements

Listing of all game records to be maintained and the responsible element.

i. Quality Assurance

Identifies responsible game elements and responsible individuals by position title.

Section III. STAFF ORGANIZATION

Gives job description and associates individual staff members with job category and position titles.

1. GAME DIRECTOR
2. QUALITY ASSURANCE
3. SYSTEMS ANALYSTS
4. PROGRAMMERS
5. CONTROL
6. PLAYER TEAMS
7. SUPPORT STAFF

Figure 1-3. (concluded)

d. Organization of Personnel:

(1) A war game consists of three distinct phases: initial preparation, production of evaluation data, and application of analytical methodologies. The employment of a single staff in all three phases demands management skill and staff flexibility. To the extent feasible, staff members should be identified early in the project with respect to their functions during dynamic play and analysis. For example, staff members who are to be Blue players should be identified early and assigned during the pregame phase to data base preparation for the Blue force structure. Then, during the analysis phase, they should be assigned to specific analysis tasks associated with the Blue forces.

(2) A representative listing of the skills required by a war game includes the following:

(a) Management:

- Game director
- Deputy director
- Technical advisor (operations research)

(b) Technical:

- Military analyst
- Operations research analyst
- Systems analyst
- Computer programmer
- Quality assurance supervisor
- Editor
- Technical assistant.

(c) Support:

- Administrator
- Keypunch operator
- Clerk/typist
- Document control clerk
- Graphic arts technician.

(3) The number of assigned individuals in each skill category will vary from game to game. For some games, two or more skills may be provided by one individual (e.g., an operations research analyst may double as a computer programmer). For other games, several individuals of similar skills will be needed to perform a particular function.

(4) Staff organization for the initial task preparation is different from that used during the production of data, but the organization should allow for an orderly transition from one phase to the other. For the entire study, the staff can be organized generally into two elements, an analysis and evaluation group and a model modification and maintenance group.

(a) During the task preparation phase the analysis and evaluation group analyzes task objectives, designs the methodology for achieving the objectives, and develops an analysis plan. This group selects measures of effectiveness and effectiveness indicators, defines performance data requirements, prepares a data base for the game, and develops a game plan. During game play and analysis, this group forms a game operations section and an analysis section, as described in succeeding paragraphs.

(b) The model modification and maintenance group is charged during the initial preparation phase with aiding in the selection of appropriate models; identifying, performing, and documenting required model modifications; assisting in data base preparation; and coding and loading input data. During production of data this group ensures a smooth interface between the game operations section and the computer. The group is responsible for liaison with the computer facility and for maintenance of tapes, decks, and disks associated with the war game model.

e. Summary. The initial preparation for a division force analysis is a critical managerial assignment. The game managers must understand the significance of the problem and determine the nature of the end product required to fulfill task objectives. They must ensure that the methodology designed for the task is complete, capable of being performed within task resources, and responsive to task requirements. They must supervise the development of a plan for analysis of data and must assemble and organize a staff representing the requisite skills for successful performance of the entire study. Satisfactory completion of the study rests obviously with thorough management planning and timely accomplishment of tasks during the preparation phase.

10. PRODUCTION OF EVALUATION DATA. The second phase of a division force analysis is instituted with the production of evaluation data. Management concern during this phase is directed toward:

- Organizing the staff for game operations
- Keeping the game on schedule
- Ensuring that requirements for evaluation data are met
- Ensuring that procedures for documenting the game are adhered to.

a. Staff Organization:

(1) The basic staff organization, consisting of the analysis and evaluation group and the model modification and maintenance group, can be retained for dynamic game play. One element within the analysis and evaluation group forms an analysis section to begin the evaluation of game-generated data according to the established analysis plan. Another element forms the team-oriented game operations section. For each game, this section forms three teams: Blue, Red, and Control. One member of each team is designated as chief. The Blue and Red teams consist of both operations and military analysts and technical assistants. The Control team is composed of operations and military analysts, programmers, and technical assistants. A mix of operations and military analysis skills in player and control teams helps to prevent methodology and computer operations problems and facilitates the solution of problems that do occur.

(2) The game is conducted according to the rules and procedures established by the game plan. Game security procedures are based on particular game requirements. If a closed game is being conducted, game intelligence is restricted, and the game rooms of the opposing team and the control team are off-limits to player team personnel.

(3) Two or more games can be conducted simultaneously to meet task objectives if personnel and facility resources permit. The USACDC War Game Facility at Fort Leavenworth, Kansas, has six fully equipped game rooms; thus, it can accommodate two closed games (with Blue, Red, and Control rooms for each) or up to six open games.

b. Game Schedule. One of the most important management concerns of the dynamic game phase is ensuring that game play progresses according to schedule. Schedule slippages during the game can result in an unacceptable compression of the time available for analysis and documentation or, alternatively, a corresponding slippage in completion of the final product. Game period turnaround time, or the calendar time required from the start of planning for one period to the start of planning for the next period, is a function of several factors, some of which are amenable to management influence.

(1) A principal factor in period turnaround time is the game-time-to-computer-time ratio achieved by the model. Model developers and the model modification and maintenance group strive for a fast running time compatible with the resolution and realism requirements of the game; however, hardware problems and system errors, including operator errors, can and do influence the game-time-to-computer-time ratio achieved in actual operation.

(2) A second principal factor in achieving a rapid game period turnaround time is the smooth and efficient operation of control and the player teams. Detailed management planning can have a significant effect on this aspect of the game operation. Procedures must exist for the tactical

planning for each period by the Blue and Red teams, the preparation of input by Control, the interface with the computer facility, the interpretation of period output by Control, and the subsequent dissemination of intelligence to player teams. These procedures must be standardized, rehearsed, and polished until they can be performed with a minimum of wasted time and motion. Situations can easily occur in game operations where A is waiting for B to complete a task, who is waiting for C to complete a task, who is waiting for A to complete a task. Detailed procedures planning is required to avoid such occurrences and to ensure that the game moves along as rapidly as possible. Volume IV, User's Manual, of the DIVWAG model documentation describes in detail dynamic play operations using the DIVWAG model. Procedures for the effective operation of player teams, control teams, and the model maintenance group are provided. Game managers will find this manual helpful in establishing procedures for dynamic play.

(3) Ideally, the calendar time allotted for the completion of a specific number of game periods or, conversely, the number of periods to be played in an allotted time, should be reasonably flexible. Game period turnaround times achievable for any specific game are extremely difficult to forecast accurately, especially when a new model or a new gaming organization is involved. Experience in running the game is often the only reliable indicator of what the lowest possible calendar-time-to-game-time ratio may be. Some form of established schedule for the production of evaluation data through gaming is inevitable, however; but study progress and completion difficulties may be avoided for management and sponsor alike by the use of conservative planning factors in the establishment of gaming schedules.

c. Evaluation Data Requirements. Management, through the analysis section, must ensure that the game produces the performance data needed to fulfill evaluation requirements. Analysts can identify areas where game-generated data are inadequate. The Control team, working within established game rules, may be able to manipulate game events to produce the required data. Alternatively, effort may be expended in the performance of side analyses, conducted outside the mainstream game, to provide data for separate analyses in areas where game data are lacking. Management's overall plan for the force analysis should provide resources for the performance of side analyses and means for their integration into the study results.

d. Documentation. Every aspect of the performance of the force analysis must be completely documented. Documentation procedures are especially important to management during the play of the game since the procedures must be rigidly adhered to, day by day, game period by game period. Game narratives are the primary vehicle for recording game events. Narratives must be prepared at the end of each game period. They may include Red and Blue status at the start of the period, plans for the period with rationale, major events of the period and their results, and the ending status of the forces. The narratives may be accompanied by graphics showing plans, the location of the FEBA, and the disposition of units.

11. APPLICATION OF THE ANALYSIS METHODOLOGY. The third phase of the division force analysis, application of the analysis methodology, can begin as soon as evaluation data are available from the game. The analysis section of the analysis and evaluation group applies the procedures established by the analysis plan to evaluate performance data by both subjective and statistical techniques. The results are documented, interpreted, and presented to the sponsor in terms of the original task objectives. Management concern during this final phase of the study effort is directed toward:

- Successful accomplishment of the analysis plan, especially the integration of the subjective and statistical analysis aspects

- Cogent presentation of the analysis results as fulfillment of the study objectives

- Final achievement of a sound basis for acceptability of results.

a. Performance of the Analysis Plan:

(1) The purpose of the analysis plan is to ensure that the analysis produces the information required by the game objectives. A research war game is played to obtain the answers to difficult questions. These questions may be very specific or very general. As an example, a specific question might concern the improvement in force performance effected by a single weapon system, and a general question might require a yes or no answer to whether Force One is better than Force Two. The analysis plan ensures that the correct data are analyzed in a form adequate to answer the questions posed by the game objectives.

(2) A war game produces an overwhelming mass of data (see Chapter 3). These data range from the subjective impressions of the staff operating the game to the straight reporting of consumption and loss figures. In satisfying the game objectives, all data, subjective and objective, must be integrated and synthesized in a manner permitting comprehension. Subjective impressions of force performance are combined with objective reporting of facts to provide credible answers to the study objectives.

b. Presentation of Analysis Results. In Paragraph 5b(7) a war game was characterized as a communication system in which participants with different reference frames were required to comprehend each other. In the presentation of analysis results this concept must be extended to include the game sponsor; he is a node in the communication system. Whatever information is obtained from the analysis must be communicated with near 100 percent comprehension. In most cases the game sponsor does not participate in the game. He is, therefore, denied the advantage of an eduction extended over a considerable period of time, an advantage enjoyed by the game participants. He has not had the opportunity to develop a common base of understanding with the

participants and manager of his game; therefore, in the presentation of results, the game manager must use the sponsor's frame of reference. Generally, considerable transformation must be anticipated; otherwise, the conclusions of the study may encounter misunderstanding, hostility, or outright disbelief. The burden is on the game manager to present results in the language of the sponsor.

c. Acceptability of Results. Management effort following the receipt of the initial force analysis directive is oriented toward constructing a sound methodology and performing a valid analytical effort as a basis for achieving a useful, responsive, scientifically and militarily acceptable product. Each aspect of the task is critical. The analysis of objectives, selection of evaluation tools, designation of measures of effectiveness, and establishment of a data base are crucial preparatory steps. The conduct of the game by sound military principles, using a well-researched and acceptable model, is fundamental. The culmination of the entire effort, however, is the analysis of data and the presentation of results. Management must ensure that the analysis procedures are visible; i.e., that the techniques are explained in detail and that their application at each step of the analysis is thoroughly documented. The methodology then is allowed to stand on its own merits, and the validity of the results derived therefrom is judged on this basis. Effective management of a successful force analysis is not confined to guiding the effort to a timely completion within project resources. Achievement of a product that reflects credit on management and sponsor alike relies on an accurate visualization of the analysis problem, design and application of an appropriate methodology, and a clear and usable presentation of the analysis results.

NOTES

CHAPTER 2

EVALUATION METHODOLOGY

Section I. INTRODUCTION

1. PURPOSE. This chapter describes a methodology for evaluating the combat effectiveness of a single force. The methodology can be used independently, when the analysis objective is single force performance, or in conjunction with the comparison methodology described in Chapter 3, when the analysis objective is the relative effectiveness of two or more forces.

2. ORGANIZATION. This introductory section provides background information relative to the philosophy guiding methodology development. Subsequent sections of this chapter present a detailed, step-by step explanation of the methodology, from the receipt of input data, through subjective and statistical procedures, to a synthesis culminating in a summary of force effectiveness. Emphasis is given to a description of the analytical procedures composing the statistical analysis steps; and an example of evaluation methodology application, using performance data from a combat simulation, is provided. The chapter concludes with a presentation of model output data arrays and a list of references applicable to the chapter.

3. BACKGROUND:

a. The evaluation methodology is intended to be applicable to data generated by any of a variety of techniques; e.g., simulations, war gaming, or field tests. It was developed independently of the DIVWAG model, but the requirements of the methodology guided the development of specifications for model output data.

b. The objective of the methodology is to provide a series of standardized processes for the evaluation of the combat effectiveness of a single force. To fulfill the objective, the methodology must provide means for:

(1) Identifying the basic objectives of a force evaluation project.

(2) Analyzing the composition of the military forces to be evaluated, the doctrine for their employment, and the probable impact of environment upon unit and system performance.

(3) Designating the appropriate measures of effectiveness (MOEs) and supporting groups of effectiveness indicators to be used.

(4) Identifying the detailed performance data required to analyze the combat activities in terms of the MOE hierarchy.

(5) Determining the scope of military activities to be simulated.

(6) Employing analytical techniques for evaluating force combat effectiveness based on performance data from simulated military operations.

c. The following paragraphs describe how the evaluation methodology incorporates the above requirements into a logical sequence of steps designed to determine the effectiveness of a single military force. For clarity, the following explanation assumes that the source of performance data is a war game and that the game is conducted using the DIVWAG model; however, the methodology may be applied to performance data derived from any appropriate source.

Section II. EVALUATION METHODOLOGY DESCRIPTION

4. GENERAL CONCEPT OF EVALUATION METHODOLOGY:

a. Discussion:

(1) The evaluation methodology comprises subjective and statistical analysis and follows a logical sequence of events to arrive at conclusions regarding the overall effectiveness of a force. The process begins with receipt of the game directive and other guidance normally provided by the proponent agency, such as any specified essential elements of evaluation.

(2) Input data from the proponent agency establishes the framework of both the evaluation effort and the war game. The force evaluation objectives are of primary importance. Their analysis provides a basis in logic for the selection of MOEs and effectiveness indicators, which, in turn, define the performance data required to support the evaluation procedures. Performance data requirements, together with other input data from the proponent agency, are used to develop the game plan. Information from the scenario, the physical characteristics of the game environment, and detailed information on the Blue and Red forces, to include necessary technical characteristics, are loaded into the computer. The game is then conducted to produce data for evaluation.

(3) The running of a single game will produce only a single data set, and statistical evaluation requires more than one set. For this reason, game output is divided into several data sets for analysis and evaluation. This division into data sets is made, first, on the basis of the mission the force is attempting to accomplish and, second, on the basis of combat activities in which subordinate units of the force are engaged individually. The force may engage in a series of combat activities, such as mobile defense, counterattack, and delay, during the conduct of the war game. The game output produced during each of these combat activities constitutes a data set for analysis. Subsets are then established to reflect subordinate unit performance by type combat activity. Statistical analysis procedures are applied to the data sets and subsets for evaluation purposes. The end products of the statistical analysis are then subjectively reviewed to identify and to assess trends and variations for military significance.

(4) After data sets and subsets have been analyzed by combat activity, they are statistically and subjectively analyzed across combat activities to evaluate the overall effectiveness of the force. If side analyses are conducted, they are correlated with the other analyses to provide the overall effectiveness summation of the force in all type combat activities.

(5) The evaluation methodology provides a comprehensive assessment of combat effectiveness based on analyses in single as well as multiple type combat activities and mission assignments. The techniques are compatible with those used in the comparison methodology (Chapter 3); thus, the evaluation methodology provides a basis for the comparative analysis of two or more forces, as well as a comprehensive assessment of the combat effectiveness of a single force of interest.

b. Procedures. The logic flow for the evaluation methodology is shown by Figure 2-1. As indicated therein, the methodology is divided into three major phases: input data, which is normally provided by the proponent agency; preparatory steps; and the performance of the evaluation.

(1) Input data include the force evaluation objectives, the scenario package, and detailed information about the forces to be gamed. The major elements of input data are described in Paragraph 5.

(2) The preparatory steps consist of actions taken by the game director and his staff in response to the evaluation directive to plan the force evaluation. These steps are described in Paragraph 6.

(3) Performance of the evaluation analysis includes the sequential steps from the beginning of the war game through the preparation of the overall force effectiveness summation. These steps are described in Paragraph 7.

5. INPUT DATA - PHASE I. Phase I consists of the receipt of input data. All input data may be provided as a single package, but more frequently they are provided on a phased basis. A possible phasing is shown below:

a. Force Evaluation Objectives. The force evaluation objectives reflect the basic questions that the proponent agency wants answered. As such, they provide guidance and direction to the entire evaluation.

b. Scenario. Normally, the scenario includes the overall missions of the opposing forces, the game environment (geographic location and time of year), and general and special situations, to include the initial deployments of major elements of the Red and Blue forces.

c. Detailed Force Compositions. Force compositions include information on the numbers and types of Red and Blue units; numbers, types, and technical characteristics of their major systems and equipment items; the details of their organizations; and the doctrine for their employment.

6. PREPARATORY STEPS - PHASE II. Six steps preparatory to the conduct of the evaluation analysis are shown on Figure 2-1. These steps are described below:

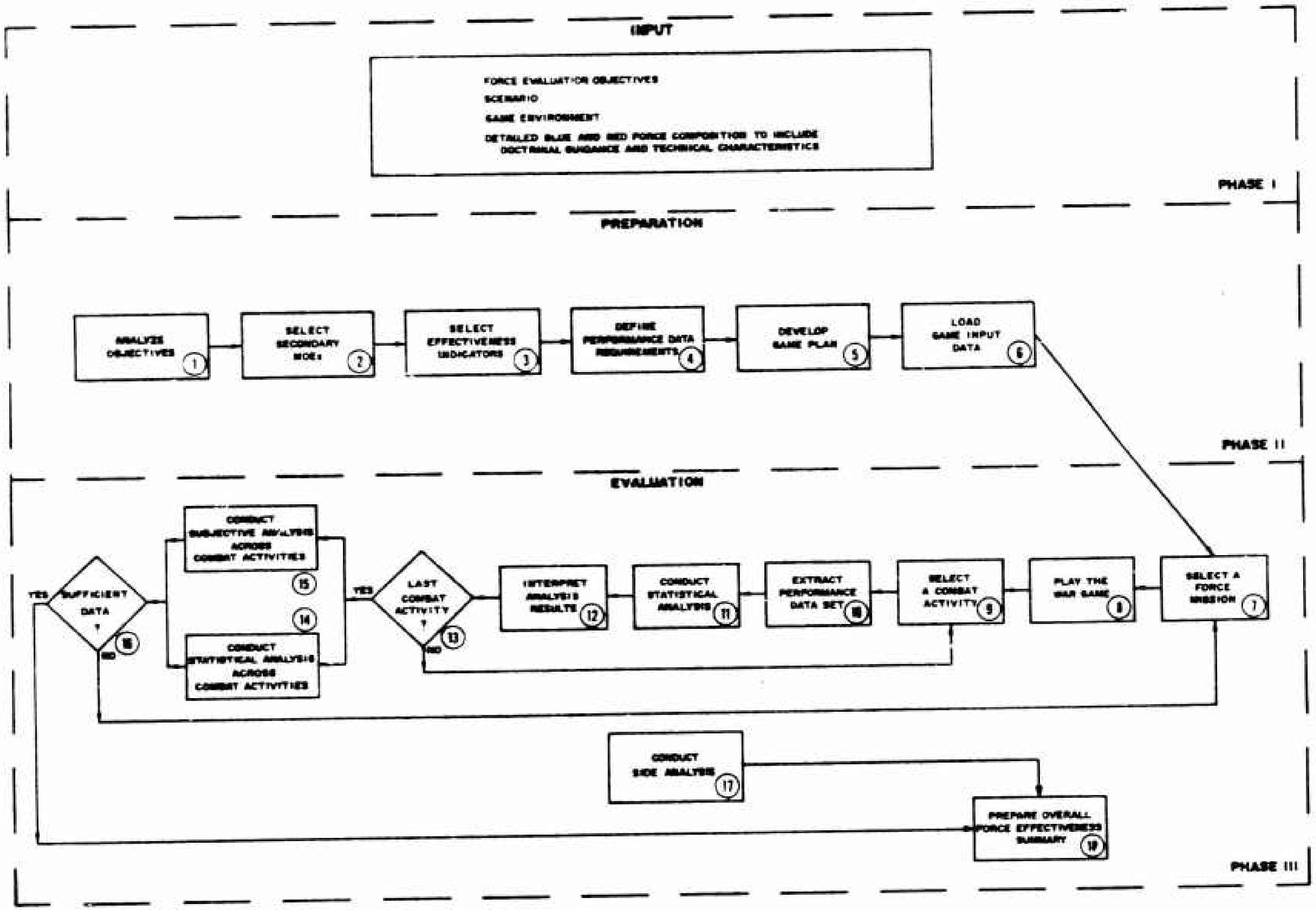


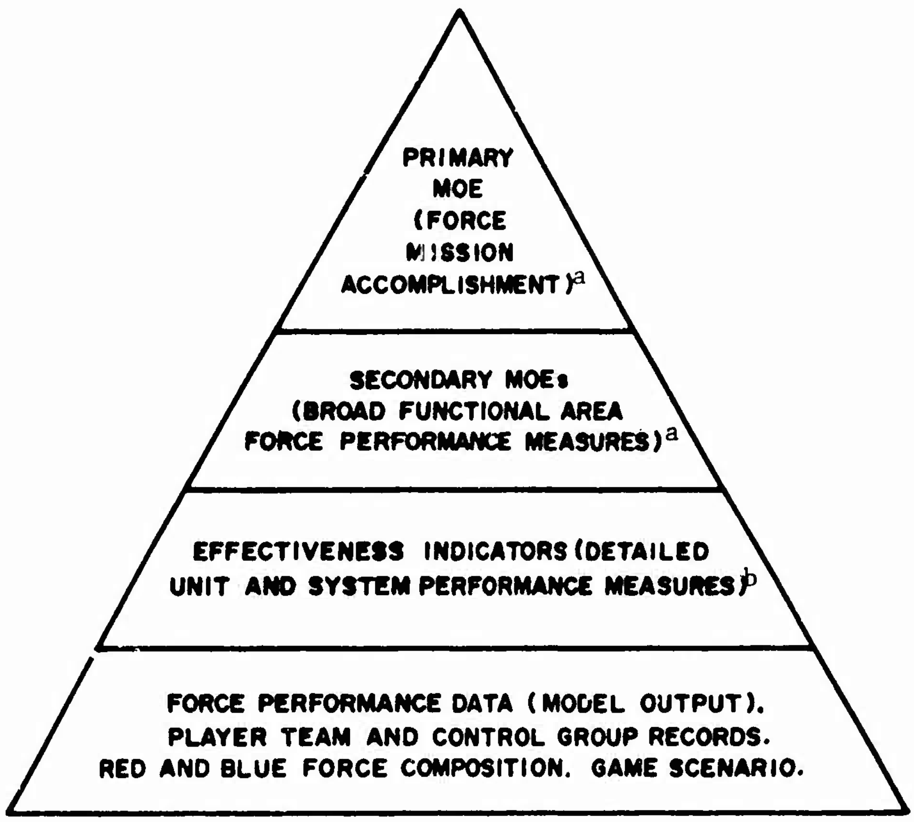
Figure 2-1. Evaluation Methodology Logic Flow

a. Analyze Objectives - Step 1. The purpose for which a study is conducted is embodied in the force evaluation objectives. The objectives must be completely stated, thoroughly understood by all involved in performing the analysis, and presented to avoid ambiguity and misinterpretation. They must be precise, specific, and leave no doubt as to the problem to be solved. The force evaluation objectives are normally provided by the proponent agency; however, they may be restated by the analysis group for clarification, if necessary, subject to approval of the proponents. Before the study proceeds, there must be unanimous agreement and understanding relative to the objectives; and the methodology must be reviewed continually in light of the objectives to ensure that the project maintains its direction.

b. Select Measures of Effectiveness and Effectiveness Indicators - Steps 2 and 3. The bases for a force evaluation are measures of effectiveness and effectiveness indicators. These measures are selected in light of the objectives of the force evaluation. In developing a methodology for evaluating force performance, CSC-CDRO established an MOE hierarchy applicable to the evaluation of force effectiveness. The pinnacle of the MOE hierarchy, as indicated in Figure 2-2, is one primary MOE, mission accomplishment; i.e., to what degree was the force successful in accomplishing its assigned mission? This primary MOE is supported in two ways:

(1) First, by four secondary measures of effectiveness, one for each of the functional areas of land combat for which meaningful quantitative output is available. (The fifth function, command, control, and communications, is not addressed discretely by the DIVWAG model at its current level of development.) These four secondary MOEs are each supported by a set of quantitative effectiveness indicators as shown by Figures 2-3 through 2-6. These secondary MOEs with their supporting effectiveness indicators provide an in-depth assessment of constituent unit and functional area performance. They serve to explain why the force succeeded or failed to the degree it did, as well as to identify force strengths and weaknesses. The effectiveness indicators supporting the four secondary MOEs will change character, and their number will expand or contract, as direct functions of the objectives of the force evaluation directive, the threat postulated, and whether those objectives require simulation of both mid and high intensity combat operation. One subset of the effectiveness indicators shown at Figures 2-3 through 2-6 is identified for use in evaluating both mid and high intensity combat simulation results; however, evaluation of high intensity combat simulation results requires use of additional effectiveness indicators. This subset is also identified at Figures 2-3 through 2-6. Effectiveness indicators are selected to highlight the most significant aspects of the performance data in light of the force evaluation objectives and the particular MOE being addressed.

(2) Second, by a series of qualitative indicators shown by Figure 2-7. This series is oriented on the stated and implied tasks included within the overall force mission. The time and terrain aspects of the mission are reflected, as are Red losses in those cases where the Blue mission specifies



a. Measure of Effectiveness - A quantitative value that indicates the degree to which a military unit or system performs its mission or achieves its goal.

b. Effectiveness Indicators - Selected quantitative elements of the large body of performance data available, chosen to highlight the most significant aspects of the performance data in light of the force evaluation objectives and the particular MOE they support.

Figure 2-2. Measure of Effectiveness Hierarchy

Secondary MOE for Intelligence Function

- Percent of available targets acquired by unit of time by area of interest.

Effectiveness Indicators for Intelligence Function

(quantified for various time periods and for as many combinations of visibility, weather, and terrain conditions and Blue and Red missions and postures as desired)

I. Applicable to mid and high intensity:

1. Daytime (nighttime) acquisitions by (1) Blue force sensor systems and by (2) intelligence sources external to the Blue force.

    a. Percent of Red aircraft within Blue area of interest[a] acquired within 2 minutes (more than 2 minutes) of arriving or becoming airborne within area.

    b. Percent of Red tank and mech rifle battalions within Blue area of interest[a] acquired within 30 minutes (more than 30 minutes) of arrival or relocation.

2. Daytime (nighttime) acquisitions by (1) Red force sensor systems and by (2) intelligence sources external to the Red force.

    a. Percent of Blue aircraft within Red area of interest[b] acquired within 2 minutes (more than 2 minutes) of arriving or becoming airborne within area.

    b. Percent of Blue maneuver battalions within Red area of interest[b] acquired within 30 minutes (more than 30 minutes) of arrival or relocation.

a. As defined in USACDC Report on Combat Commanders' Surveillance and Target Acquisition Needs (1969-1975) dated October 1969 for battalion, brigade, and division.

b. Assumed to be same size as Blue's.

Figure 2-3. Secondary MOE and Effectiveness Indicators for Intelligence Function (continued next page)

II. Applicable to high intensity only:

1. Daytime (nighttime) acquisitions by (1) Blue force sensor systems and by (2) intelligence sources external to the Blue force.

   - Percent of Red nuclear capable delivery systems, other than aircraft, within Blue area of interest[a] acquired within 30 minutes (more than 30 minutes) of arrival or relocation.

2. Daytime (nighttime) acquisitions by (1) Red force sensor systems and by (2) intelligence sources external to the Red force.

   - Percent of Blue nuclear capable delivery systems, other than aircraft, within Red area of interest[b] acquired within 30 minutes (more than 30 minutes) of arrival or relocation.

a. As defined in USACDC Report on Combat Commanders' Surveillance and Target Acquisition Needs (1969-1975) dated October 1969 for battalion, brigade, and division.

b. Assumed to be same size as Blue's.

Figure 2-3. Secondary MOE and Effectiveness Indicators for Intelligence Function (concluded)

Secondary MOE for Firepower Function

- Personnel casualties and key equipment losses inflicted per unit of time as a percent of the total force.

Effectiveness Indicators for Firepower Function

(quantified at end of each significant event such as reinforcement (change in force ratio) or termination of battle and for as many combinations of visibility, weather, and terrain conditions and Blue and Red missions and postures as desired)

I. Applicable to mid and high intensity:

1. Average hourly percent of all Red (Blue) personnel that became casualties as a result of Blue (Red) firepower.

2. Average hourly percent of total Red (Blue) personnel casualties caused by (1) Blue (Red) ground based direct fire weapons and by (2) Blue (Red) field artillery (including aerial field artillery).

3. Average hourly percent of all Red (Blue) tanks damaged or destroyed by Blue (Red) firepower.

4. Average hourly percent of total Red (Blue) tank losses caused by (1) Blue (Red) ground based direct fire weapons, by (2) Blue (Red) attack helicopters, and by (3) Blue (Red) close air support.

5. Average hourly percent of all Red (Blue) attack helicopters damaged or destroyed by Blue (Red) firepower.

6. Percent of total Red (Blue) attack helicopter losses (quantified for each attack helicopter mission) caused by (1) Blue (Red) ground based air defense systems and by (2) Blue (Red) aircraft.

7. Average hourly percent of all Red (Blue) air defense fire units destroyed by Blue (Red) firepower.

8. Average hourly percent of all Red (Blue) air defense fire unit losses caused by Blue (Red) (1) ground based direct fire weapons, (2) field artillery (including aerial field artillery), (3) CAS, and (4) attack helicopters.

Figure 2-4. Secondary MOE and Effectiveness Indicators for Firepower Function (continued next page)

9. Ratio of Blue (Red) average hourly loss rates of personnel and tanks to Red (Blue) equivalent average hourly loss rates:*

$$B_c/R_c\ ,\quad B_t/R_t\ ,\quad \frac{B_c/B_o}{R_c/R_o}\ ,\quad \frac{B_t/B_o}{R_t/R_o}$$

10. Ratio of Blue (Red) attack helicopter losses from all types of Red (Blue) firepower to the number of Red (Blue) tank losses attributed to Blue (Red) attack helicopters:

$$B_h/R_t\ ,\quad \frac{B_h/B_o}{R_t/R_o}\quad \text{and}\quad R_h/B_t\ ,\quad \frac{R_h/R_o}{B_t/B_o}$$

11. Ratio of the Red (Blue) personnel casualty rate from all causes to the equivalent Blue (Red) personnel casualty rate during that hour in which the Red (Blue) personnel casualty rate was at a maximum:

$$\frac{\text{Max }(R_c)}{B_c}\quad \text{and}\quad \frac{\text{Max }(B_c)}{R_c}$$

12. Ratio of the Red (Blue) tank loss rate from all causes to the equivalent Blue (Red) tank loss rate during that hour in which the Red (Blue) tank loss rate was at a maximum:

$$\frac{\text{Max }(R_t)}{B_t}\quad \text{and}\quad \frac{\text{Max }(B_t)}{R_t}$$

13. Ratio of the number of Blue (Red) attack helicopter losses from all causes to the number of Red (Blue) tank losses attributed to Blue (Red) attack helicopters during that hour in which Blue (Red) attack helicopter losses were at a maximum:

$$\frac{\text{Max }(B_h)}{R_t}\quad \text{and}\quad \frac{\text{Max }(R_h)}{B_t}$$

* See note at end of figure.

Figure 2-4. Secondary MOE and Effectiveness Indicators for Firepower Function (continued)

14. Ratio of the number of Blue (Red) attack helicopter losses from all causes to the number of Red (Blue) tank losses attributed to Blue (Red) attack helicopters during that hour in which Red (Blue) tank losses attributed to Blue (Red) attack helicopters were at a maximum:

$$\frac{B_h}{\text{Max }(R_t)} \quad \text{and} \quad \frac{R_h}{\text{Max }(B_t)}$$

II. Applicable to high intensity only:

1. Percent of total Red (Blue) personnel casualties caused by (1) Blue (Red) field artillery nuclear rounds and by (2) Blue (Red) close air support nuclear bombs.

2. Percent of total Red (Blue) tank losses caused by (1) Blue (Red) field artillery nuclear rounds and by (2) Blue (Red) close air support nuclear bombs.

3. Percent of total Blue (Red) attack helicopter losses caused by (1) Red (Blue) field artillery nuclear rounds and by (2) Red (Blue) nuclear bombs.

Note: B_c = Blue personnel casualties, R_c = Red personnel casualties, B_t = Blue tank losses, R_t = Red tank losses, B_h = Blue attack helicopter losses, R_h = Red attack helicopter losses, B_o = Blue organization total, R_o = Red organization total.

Figure 2-4. Secondary MOE and Effectiveness Indicators for Firepower Function (concluded)

Secondary MOE for Mobility Function

- Average speed (kilometers per hour) of unit movements.

Effectiveness Indicators for Mobility Function

(quantified for as many combinations of visibility, weather, and terrain conditions and Blue and Red missions and postures as desired)

I. Applicable to mid and high intensity:

1. Average speed (kilometers per hour) of all ground tactical movements* by Blue (Red):

   a. Maneuver battalions (including air and ground cavalry).

   b. Field artillery battalions.

   c. Engineer battalions.

2. Average speed (kilometers per hour) of all airmobile tactical movements* by Blue (Red):

   a. Maneuver battalions (including air and ground cavalry).

   b. Field artillery battalions.

   c. Engineer battalions.

3. Average speed (kilometers per hour) of all tactical movements* by Blue (Red):

   a. Maneuver battalions (including air and ground cavalry).

   b. Field artillery battalions.

   c. Engineer battalions.

II. Applicable to high intensity only:

None.

* As defined in AR 320-5.

Figure 2-5. Secondary MOE and Effectiveness Indicators for Mobility Function

Secondary MOE for Combat Service Support Function

- Shortages as a percent of total requirement.

Effectiveness Indicators for Combat Service Support Function

(quantified for various time periods and for as many combinations of visibility, weather, and terrain conditions and Blue and Red missions and postures as desired)

I. Applicable to mid and high intensity:

1. Average Blue (Red) Class III unfulfilled daily resupply requirement expressed as a percent of the total requirement.

2. Average Blue (Red) Class V unfulfilled daily resupply requirement expressed as a percent of the total requirement.

II. Applicable to high intensity only:

None.

Figure 2-6. Secondary MOE and Effectiveness Indicators for Combat Service Support Function

Qualitative Indicators for Use in Determining the Degree of Mission Accomplishment

(For any given mission, the qualitative indicators used to identify the degree of mission accomplishment will vary; however, the following will always apply.)

1. Time to execute mission as a function of time required.

2. The gain (loss) of key terrain as a function of mission.

3. Total number and hourly rate of Red (Blue) personnel casualties, tank losses, and attack helicopter losses in those cases where the Blue (Red) mission specified attrition or destruction of enemy forces. (See MOE and effectiveness indicators for firepower function.)

4. Residual status of Blue (Red) maneuver, field artillery, and attack helicopter battalions:

   a. Personnel strength as a percent of authorized.

   b. Key equipment item strength as a percent of authorized.

Figure 2-7. Qualitative Indicators for Use in the Subjective Analysis of Mission Accomplishment

attrition or destruction of enemy forces. (Red losses are also considered in all cases under the secondary MOE for the function of firepower as shown at Figure 2-4.) The residual status of key portions of the force is also reflected here.

c. Define Performance Data Requirements - Step 4. Once the force evaluation objectives and the MOEs with their supporting effectiveness indicators are identified, it is possible to determine specific performance data requirements. The organization of available DIVWAG model output data is arrayed by Figure 2-8. As shown by that figure, model output in the form of Blue and Red performance data is organized by force mission for selected type subordinate units and functional areas. Section V of this chapter contains a set of arrays showing the organization of model output data for each of the major categories marked by an asterisk in Figure 2-8. Section V also contains an explanation, with examples, of the use of the model output data arrays.

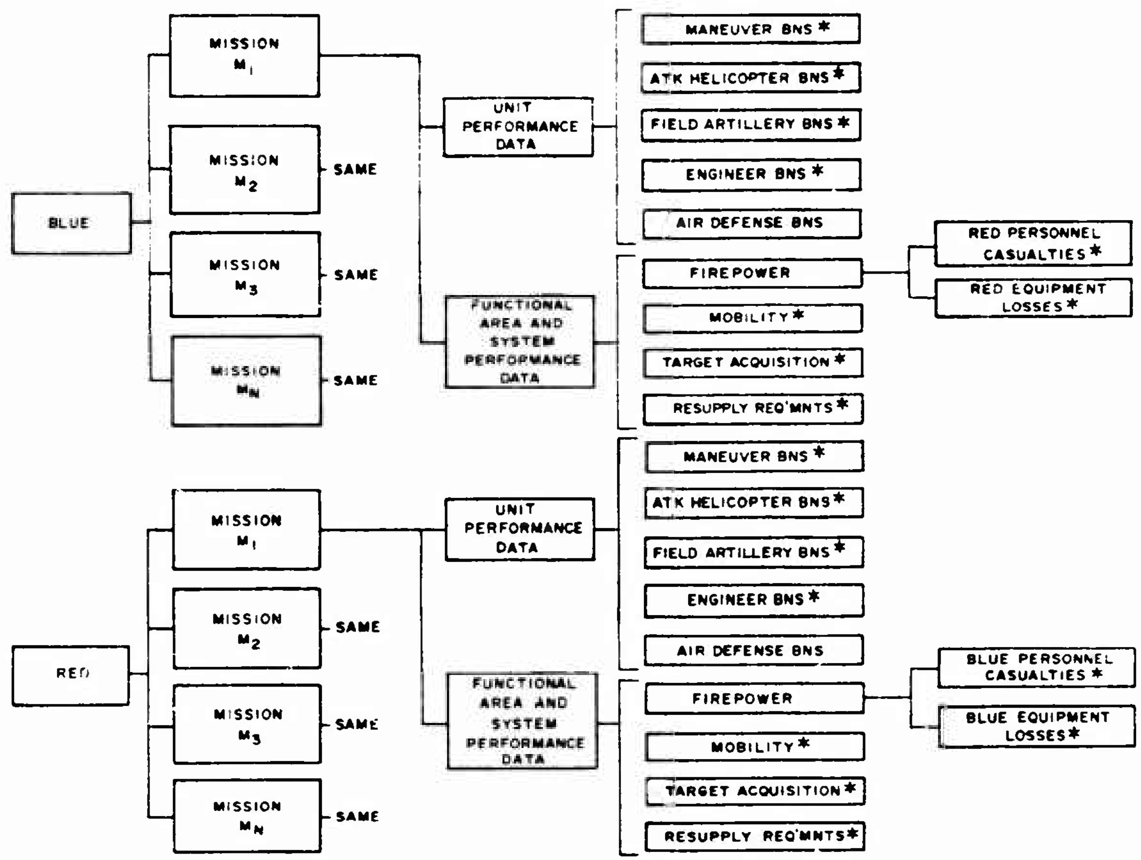
d. Develop Game Plan - Step 5. The game plan is the director's detailed program for utilization of available resources (computer hours, physical facilities, and skilled manpower) to execute the war game.

e. Load Game Input Data - Step 6. This step consists of inserting into the computer system the detailed game input data. Included are the terrain characteristics; weather data for the appropriate location and time of year; the number, type, initial locations, and performance characteristics of Red and Blue forces; and the types, amounts, and technical and performance characteristics of their equipment.

7. CONDUCT OF EVALUATION ANALYSIS - PHASE III. Performance of the evaluation analysis comprises 12 steps (Steps 7 through 18 of the evaluation methodology logic flow, Figure 2-1) as described below.

a. Select a Force Mission - Step 7. When game input data have been loaded, a force mission is selected in consonance with the game scenario; and Blue and Red initial situation intelligence summaries and operation orders are prepared. These actions represent the initiation of dynamic play, and available resources of computer time and skilled manpower are applied in accordance with the game plan.

b. Play the Game - Step 8. This step includes actions by the control group and the player teams as well as application of the model. As stated in paragraph 7a., action by the control group, in the form of issuing the initial Blue and Red intelligence summaries, normally marks the beginning of dynamic play. Next, Blue and Red operations orders are prepared, encoded, and inserted into the computer system. Exercise of the model produces two forms of output. One is a report back to the control group. This report provides information on what happened to the Blue and Red units as well as their current location and status. The control group uses this report to initiate the next cycle of dynamic play. The second form of model output is detailed performance data used to feed the evaluation processes.



* - SEE FIGURES 2-25 THROUGH 2-44.

Figure 2-8. Organization of Model Output Data

c. Select a Combat Activity - Step 9. Evaluation of the performance of subordinate units of the Blue force necessitates identification of the combat activities in which they participated individually during the time the force as a whole was attempting to accomplish a force mission. The types of such combat activities, their duration, and the background conditions that existed; e.g., visibility, weather, terrain, and posture, provide the basis for corresponding subsets of model output data. At Step 9, one such combat activity is selected for analysis.

d. Extract Performance Data Set - Step 10. Dynamic play produces performance data in the form of game records and model output. The model provides quantitative performance data output. The game records prepared by the Red and Blue player teams and the control group provide a history of the period-to-period missions assigned to the subordinate elements of the Red and Blue forces, the tactical plans for the accomplishment of those missions, and the unit after-action reports describing the simulated operations. Game records are critical sources of information for use in the evaluation process, and they serve to place the model output in perspective. Performance data packages f model output for the combat activity selected at Step 9 are extracted, together with the supporting package of game records. Both types of data are subjected to quality control examination prior to initiation of any force evaluation processes.

e. Conduct Statistical Analysis - Step 11. A statistical analysis of selected effectiveness indicators using data from the simulation of a single combat activity is now conducted. The statistical procedures associated with this analysis are outlined in Section III of this chapter.

f. Interpret Analysis Results - Step 12. The results of the statistical analysis are reviewed by military analysts to determine why the units performed as they did and to evaluate the military significance of any trends and variations detected. This review includes an examination of the battlefield background conditions, the missions assigned to the subordinate units being analyzed, and the orders given to them for the performance of those missions. Additionally, this review provides an opportunity for the development of insights into the reasons for deficiencies in mission accomplishment. (Insights are defined as intuitive observations not completely supported by available game data.) The review culminates in a summary statement of how the subordinate units performed in the particular combat activity analyzed and why they performed as they did.

g. Last Combat Activity - Step 13. At this point in the evaluation a check is made to determine if all identified combat activities have been analyzed. If not, Steps 9 through 12 are repeated until the data subset of each subordinate unit combat activity identified within the overall force mission has been analyzed.

h. Conduct Statistical Analysis Across Combat Activities - Step 14. A statistical analysis of selected effectiveness indicators using data from the simulation of two or more combat activities is conducted to support the subjective analysis described in Step 15. The statistical procedures associated with this analysis are outlined in Section III of this chapter.

i. Conduct Subjective Analysis Across Combat Activities - Step 15. After all type combat activities have been individually analyzed, an analysis of force performance across combat activities is conducted. This analysis draws on the results of the statistical analysis across combat activities (Step 14) and on the previously prepared summaries of subordinate unit performance by combat activity (Step 12). It considers data required by the MOEs and effectiveness indicators listed at Figures 2-3 through 2-7 as well as all other game records and model output. This analytical effort culminates in a summary description of mission accomplishment by the force. The summary includes:

(1) A statement of the mission to include both stated and implied tasks.

(2) A statement of whether the mission was accomplished.

(3) Identification of the degree of success achieved.

(4) A listing of the identified strengths and weaknesses with an assessment of their significance:

(a) Where the strengths and weaknesses are peculiar to a single combat activity.

(b) Where the strengths and weaknesses extend across two or more combat activities.

(5) A comparison of all insights derived from the evaluation of subordinate units by combat activity, with consistencies and inconsistencies identified.

j. Sufficient Data - Step 16. At this point in the evaluation a check is made to determine if the evaluation of the force completed in Step 15 is an adequate response to the force evaluation objectives. If not, Steps 7 through 15 are repeated for additional force missions until the performance data available will support an adequate force evaluation.

k. Perform Side Analysis - Step 17. A side analysis is an evaluation of game events conducted separately from the main flow of the game. The scope and magnitude of a side analysis can vary, and the analytical method used is dependent on the problem. One method is a parametric analysis, in which an existing submodel, a modified submodel, or a special submodel is exercised repeatedly to review a set of war game conditions, varying one cond' ion at

a time to measure the resultant change. Another method is a detailed study of a subject derived from the war game situation but not addressed in the dynamic play of the game. For example, if the dynamic play of the game did not simulate the operations of medical units, a study could be made to determine the adequacy of such units within the force being gamed under the various situations developed in dynamic play. The evaluation methodology can accommodate any number of such side analyses, subject to the availability of time and resources.

l. Prepare Overall Force Effectiveness Summary - Step 18. The final step of the force evaluation methodology is preparation of an overall force effectiveness summary. This summation draws on the summaries previously prepared for force performance of individual missions (Step 15); on the results of all statistical analyses (Steps 11 and 14); and on the results of any side analyses (Step 17) conducted. The summary is oriented to responding to the force evaluation objectives. It reflects the degree of mission accomplishment attained by the force, the identified strengths and weaknesses of the force, and any insights secured as to the reasons for deficiencies in mission accomplishment.

Section III. STATISTICAL ANALYSIS

8. INTRODUCTION. This section presents the details of the statistical analysis procedures applied in the force evaluation methodology. Section II indicated that statistical analysis is used first to evaluate subordinate unit performance during a single type combat activity (Step 11) and then to assess subordinate unit performance across all combat activities simulated (Step 15). The following paragraphs describe the statistical techniques used in the two steps and the rationale for their selection. Section IV then illustrates their use with performance data generated by an actual simulation process.

9. GAME DATA ARRAYS:

a. Each unit participating in the simulated combat is a source of data that can be analyzed. Statistical analysis applied to the data can help determine force strengths and weaknesses. This must be done as a function of background conditions of varying terrain, weather, vegetation, and engagements/battles. (The term "background conditions" corresponds to the statistical term "treatments," which is used extensively in the following discussion.) The first step is to place the data into proper subsets.

b. The totality of data from a combat simulation can be visualized using Figure 2-9,

where: i = unit type

j = effectiveness indicator type

k = activity type

ijk represents the three types.

c. Each subscripted block contains engagement data for a fixed unit type, a fixed activity type, and a fixed effectiveness indicator. Since these data are collected for a variety of activities over an extended period of time, the data collected for each block can be placed into subsets. Each subset of data is to be a collection of information over a fixed set of background conditions; e.g., terrain, brightness, vegetation, weather, and other significant factors that influence detection, recognition, and combat activity in general. When the data are placed in their proper subsets in this manner, there will be a row of engagement data for each subset or treatment type. Thus, each triple subscript corresponds to a collection of engagement data presented in a two-dimensional array.

d. For each treatment type the final ranks, developed through the use of statistical testing, allow the analyst to draw inferences about unit (or weapon system) performance. From this set of ranked arrays it is possible to determine which of the treatments was most significant in the total unit

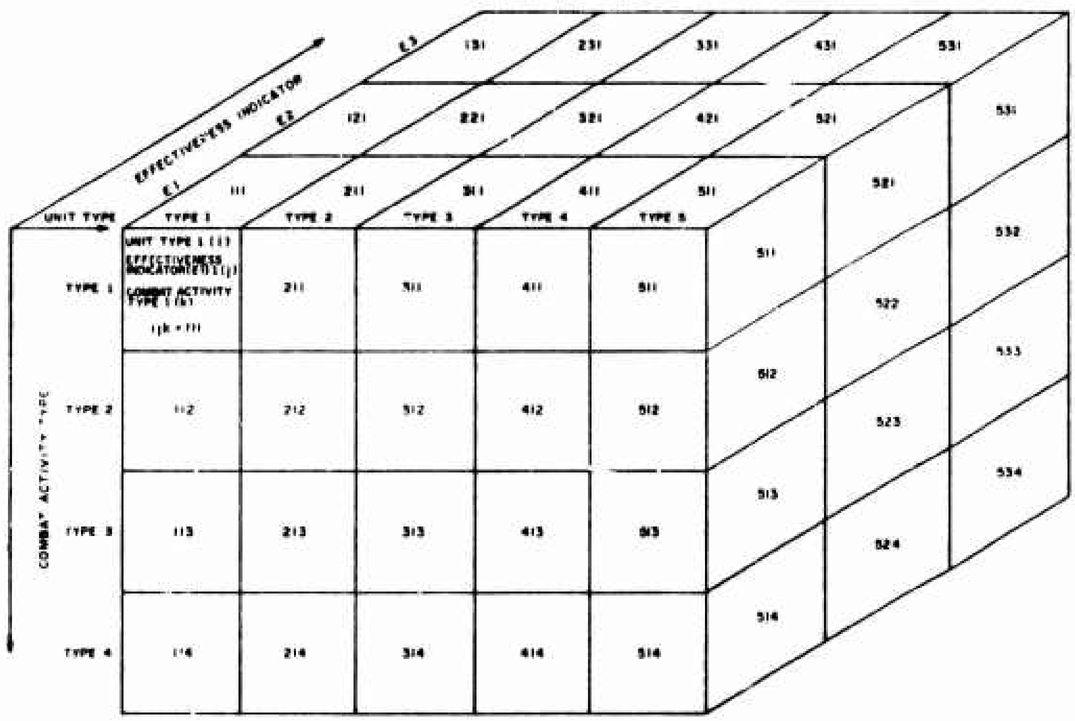


Figure 2-9. Data Array

performance. Evidently, if one treatment is consistently ranked above the others, this treatment must be considered in detail, since it "drove" the game output. This fact can be incorporated in the subjective analysis, which is a necessary complement to the statistical analysis.

10. STATISTICAL TECHNIQUES. After the data are arrayed as shown in Figure 2-9, statistical analysis procedures can be applied. A series of well-defined statistical analysis techniques is incorporated in the evaluation methodology. These techniques are designed to provide an objective basis for the identification of significant strengths and weaknesses of division force elements and models. Although most of these tests and concepts are well known and are treated in many texts (references 1, 2, 3, 4, and 5), this paragraph provides a brief description of each as background to aid the reader in understanding the statistical application described in subsequent paragraphs.

a. Statistical Hypothesis:

(1) A statistical hypothesis is an assertion about the distribution of one or more random variables. If the statistical hypothesis completely specifies the distribution, it is called a simple statistical hypothesis; if it does not, it is called a composite statistical hypothesis.

(2) A null hypothesis (H_0) is the statistical hypothesis that is subjected to a test. The hypothesis that remains tenable if the null hypothesis is rejected is called the alternative hypothesis (H_1). Hypothesis testing can be viewed as a procedure whereby an experimenter decides which one of a dichotomous set of mutually exclusive and exhaustive hypotheses is to be rejected and which one is to be accepted at some specified risk of making an incorrect decision.

(3) In the selection of forces for evaluation in a project, the military planner has in mind a hypothesis (or several hypotheses) with regard to force effectiveness. In conducting the evaluation, the analysts state a hypothesis (or sets of hypotheses) that must be satisfied yet can be tested only after running the model. Typical of the null hypotheses that may be tested are those shown in Figure 2-10.

(4) In both situations shown in Figure 2-10 there is stated a hypothesis regarding force effectiveness, which will be either accepted or rejected on the basis of model outputs. When a hypothesis is rejected or accepted on the basis of circumstances or special gamer developed tactics, the fact that human intervention played an important role in force effectiveness can be noted. Likewise, proper measures of force effectiveness that are solely weapons mix dependent can be used to test the null hypothesis H_0: There is no difference in unit performance as weapons mix is varied.

SINGLE COMBAT ACTIVITY

1. Battle characteristic$_i$[1] was not a statistically significant variable in determining the relative ranking[2] achieved by unit$_j$ across all measures of effectiveness (MOEs) and/or effectiveness indicators (EIs).

2. There is no statistically significant difference in the relative rankings achieved by units equipped with weapon system$_i$, as compared to the rankings achieved by similar type units equipped with weapon system$_j$, across all MOEs and/or EIs.

3. There is no statistically significant difference in the relative rankings achieved by units equipped with sensor mix$_i$, as compared to the rankings achieved by similar type units equipped with sensor mix$_j$, across all MOEs or EIs.

TWO OR MORE COMBAT ACTIVITIES

1. Battle characteristic$_i$ was not a statistically significant variable in determining the relative rankings achieved by unit$_j$ across all MOEs and combat activities[3].

2. There is no statistically significant difference in the relative rankings achieved by units equipped with weapon system$_i$, as compared to the rankings achieved by similar type units equipped with weapon system$_j$, across all MOEs and combat activities.

3. There is no statistically significant difference in the relative rankings achieved by units equipped with sensor mix$_i$, as compared to the rankings achieved by similar type units equipped with sensor mix$_j$, across all MOEs and combat activities.

4. There is no statistically significant difference in the relative rankings for MOE$_i$ achieved by unit$_j$ across all combat activities.

5. There is no statistically significant difference in the relative rankings for MOE$_i$ achieved by the division force across all combat activities.

Notes:

[1] Example battle characteristics consist of the following:

Visibility condition	Duration of the battle
Terrain type	Force ratio
Blue mission	Blue weapons mix
Red mission	Red weapons mix
Blue posture	Blue sensor mix
Red posture	Red sensor mix

[2] The term "relative rankings" refers to the cardinal number associated with the comparative performance of different type units within a major category, as measured by a particular MOE; e.g., different types of maneuver units, field artillery units, or DAFS units.

[3] Combat activities include attack/counterattack, defense, withdraw/delay, and covering force.

Figure 2-10. Typical Null Hypotheses

Testing these hypotheses is equivalent to asking a question about the system of interest, rephrasing the question so only one of two possibilities is true, and then testing to determine which is correct.

(5) In any case the planner and the analyst work together in that they ask the same questions, which can be answered only by the game data output. Both planner and analyst are interested in developing a good model so the proper inference can be made when the null hypothesis is either accepted or rejected. The procedure by which this decision is made is called a statistical test. The next paragraph deals with the type of statistical tests that can be employed to determine force effectiveness using output data from a war game or set of simulations constituting an entire game.

b. Statistical Test. A statistical test is the comparison of two hypotheses in the light of sample data according to a set of decision rules. A test of a statistical hypothesis is a rule which, when the experimental sample values have been obtained, leads to a decision to accept or to reject the hypothesis under consideration. A variety of standard statistical tests are well known and can be used in hypothesis testing; however, in some research situations, it is not possible to specify the functional form of the population distribution as is necessary in parametric statistics. Statistical procedures that do not depend on a knowledge of population distributions and associated parameters are called nonparametric or distribution-free methods. These nonparametric methods are used when the researcher is generally unable or unwilling to assume that the underlying populations are normally (Gaussian) distributed or have equal variances (homoscedastic). The nonparametric methods have been selected for use in the force evaluation methodology. Most prominent among these tests are:

. The Kruskal-Wallis Test

. The Friedman Test

. The Mann-Whitney Test

(1) Kruskal-Wallis One-Way Analysis of Variance (ANOVA) by Ranks. Kruskal and Wallis (1952) developed a nonparametric test based on ranks. Their one-way analysis of variance by ranks provides a test of the null hypothesis that k independent samples were drawn from k identically distributed populations. It is assumed that the data provide at least ordinal information and that the underlying probability distribution is continuous. The Kruskal-Wallis one-way ANOVA by ranks proceeds as follows:

(a) Rank the N items from largest to smallest. Should ties occur, the tied items are ranked according to the mean rank of all items in the tie group. (The reliability of this test remains unaltered if items are ranked from smallest to largest.)

(b) Count the number in each tie group and call this number t_i. Calculate $T_i = t_i^3 - t_i$ for each tie group.

(c) Arrange the rankings into columns by mission and calculate the column rank sum R_i.

(d) Calculate:

$$H = \left[\frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^{\,2}}{n_i} - 3(N+1) \right] / 1 - \frac{\sum T_i}{N^3 - N}$$

where:

k = number of rank sums; i.e., number of columns

n_i = number of entries in each column

N = total number of samples

(e) Reject the null hypothesis (H_0) at significance level α if $H > \chi^2_\alpha$ with k-1 degrees of freedom.

(2) Friedman Two-Way Analysis of Variance (ANOVA) by Ranks. Friedman (1937) has developed a nonparametric test based on ranks that can be used when matched subjects are obtained. Friedman's two-way analysis of variance by ranks provides a test of the null hypothesis that k related samples were drawn from k identically distributed populations. The Friedman two-way ANOVA proceeds as follows:

(a) Cast the scores into a two-way table having N columns (conditions) and k rows (subjects or groups).

(b) Rank the scores in each column from 1 to k.

(c) Determine the sum of ranks in each row R_j.

(d) Compute:

$$\chi_r^2 = \frac{12}{Nk(k+1)} \sum_{1}^{k} (R_j)^2 - 3N(k+1)$$

(e) Use tables (Seigel) to determine probability of occurrence of null hypothesis. For large N and/or k, χ_r^2 is equivalent to the parametric chi-square test with k-1 degrees of freedom.

(f) If the probability yielded by the appropriate method is equal to or less than α reject the null hypothesis H_0.

(3) Mann-Whitney U-Test:

(a) The Mann-Whitney U-statistic can be employed to make orthogonal comparisons among the k treatment population or to determine which comparisons among the k treatment population are significant. In conjunction with the U-tests, a pairwise analysis must be employed in order to assign final ordinal ranks to the population.

(b) The general scheme of pairwise testing using the Mann-Whitney U-Test should be as follows:

1. Test pairs k-1 units apart.

2. Test pairs k-2 units apart.

3. Test pairs k-3 units apart.

. .

. .

. .

k-1. Test pairs 1 unit apart.

4. When no difference is noted between two pairs, assign an average rank to those in the set between, and including, the two options tested and proceed to whichever set of pairs contains only the remaining untested options.

(c) The Mann Whitney U-Test then proceeds stepwise as follows:

1. After determining the size of each of the two groups being tested, call the number of items in the smallest group n_1 and the number of items in the other group n_2.

2. Rank the $n_1 + n_2$ items from 1 to $n_1 + n_2$ by assigning the smallest item the rank 1. (This could be done in reverse order and the largest item would then receive a rank 1).

3. Compute:

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

where: R_1 = rank sum for the group n_1

R_2 = rank sum for the group n_2

4. Choose U to be the smallest of U_1 and U_2.

5. Assuming $n_2 \geq 20$ calculate the normalized value:

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2}{(N-1)} \left(\frac{N^3 - N}{12} - \sum T_i \right)}}$$

(For smaller sample size calculations see Siegel (reference 1).)

where:

$N = n_1 + n_2$

$T_i = t_i^3 - t_i$ and t_i = number of elements in i^{th} tie group.

6. Use the value Z to test the null hypothesis of no difference between options. The test is exactly the same as the normal distribution test of the hypothesis that the means of two samples are drawn from the same population.

(4) Nemenyi Test:

(a) An alternative a posteriori procedure for determining which pairwise comparisons among k treatment populations are significant was proposed by Nemenyi (1963). This procedure involves less computational labor than the Mann-Whitney U-Test and is based on the Kruskal-Wallis test. In order to reject the hypothesis that two samples j and j′ were drawn from identically distributed populations the absolute value of the difference d for ranks j and j′ must exceed dKW where:

$$d = \left| \left(\frac{1}{n_j} \sum_{i=1}^{n} R_{ij} \right) - \left(\frac{1}{n_{j'}} \sum_{i=1}^{n} R_{ij'} \right) \right|$$

and:

$$dKW = \sqrt{H_{\alpha, k-1}} \cdot \sqrt{\frac{N(N+1)}{12}} \cdot \sqrt{\frac{1}{n_j} + \frac{1}{n_{j'}}}$$

The Kruskal-Wallis test requires that data be arrayed, ranks applied, and a quantity H calculated. When pairwise testing, d represents the difference between mean rank for the two rows corresponding to the two conditions under test; H_α, k-1 represents that value which H, calculated by the Kruskal-Wallis test, must exceed in order for the Kruskal-Wallis test to reject the null hypothesis; N is the total population under test; and the constants n_j and $n_{j'}$ represent the number of entries in rows j and j'.

(b) By analogy, the two-way test requires that the absolute difference d for ranks j and j' exceed dF, where:

$$d = \left|\left(\frac{1}{n} \sum_{i=1}^{n} R_{ij}\right) - \left(\frac{1}{n} \sum_{i=1}^{n} R_{ij'}\right)\right|$$

and:

$$dF = \sqrt{\chi^2_{r,\alpha}} \cdot \sqrt{\frac{k(k+1)}{6n}}$$

In these expressions, n represents the number of ranked columns, k is the number of rows in the array to be tested, and $\chi^2_{r,\alpha}$ is equivalent to χ^2_α, k-1.

(c) These tests have the advantage that throughout their execution the significance level α is maintained for every test. Also, they require that the input data be ranked only once, thus decreasing computer time. The treatment of game output data can proceed using either the Mann-Whitney U-test or the Nemenyi test. For purposes of illustration the Mann-Whitney U-test has been incorporated into the procedures described by this section.

c. Hypothesis Testing. Hypothesis-testing procedures are tools that aid an experimenter in interpreting the outcome of research. Such procedures should not be permitted to replace the judicial use of logic by an alert analytic experimenter. In particular, the technique of nonparametric analysis of variance described above should be considered an aid in summarizing data. It should be used to help an analyst understand what went on during conduct of the war game so that proper inferences can be made to assist in the decision making process.

d. Significance Level:

(1) For every null hypothesis H_0, there exists at least one alternative hypothesis H_1. An a priori procedure is to reject H_0 in favor of H_1 if a statistical test yields a value whose associated probability of occurrence under H_0 is equal to or less than some small probability sympolized as α. That small probability is called the level of significance. There are two types of errors which may be made in arriving at a decision about H_0.

(a) A Type I error is the rejection of H_0 when in fact the hypothesis is true. The significance level α is the probability that a statistical test will yield a value for which the null hypothesis will be rejected when in fact it is true. That is, the significance level indicates the probability of committing the Type I error.

(b) A Type II error is the acceptance of H_0 when in fact the hypothesis is false. β is the probability that a statistical test will yield a value for which the null hypothesis will be accepted when in fact it is false. That is, β gives the probability of committing the Type II error.

(2) The power of a test, $1 - \beta$, tells the probability of rejecting the null hypothesis when it is false (and thus should be rejected). Power is related to the nature of the statistical test chosen and depends upon the sample size N.

11. CONCEPT FOR STATISTICAL ANALYSIS IN SINGLE FORCE EVALUATION METHODOLOGY:

a. Introduction. The statistical analysis within the evaluation methodology is an approach to determining significant strengths and weaknesses of a force structure. It can also be used to aid the model builder by indicating those models or portions of a model that are insensitive to variations in battle characteristics/combat conditions (or treatments, in the statistical sense); therefore, the techniques described herein should be viewed also as an integral part of the experimental design of a model. Experimental design refers to five interrelated activities required in the investigation of scientific or research hypotheses. These activities, listed in the order performed, are as follows:

(1) Formulate statistical hypotheses and make plans for the collection and analysis of data to test the hypotheses.

(2) State decision rules to be followed in testing the hypotheses.

(3) Collect data according to plan.

(4) Analyze data according to plan.

(5) Make decisions concerning the statistical hypotheses based on decision rules and inductive inferences concerning the probable truth or falsity of research hypotheses.

b. Procedures:

(1) The objective of the statistical analysis is to present a coherent and logical mathematical basis for the analysis of the effectiveness of a single force of division size. In this context comparisons are made among constituents of the division.

(2) The output from the conduct of a war game presents a myriad of data to the analyst. This data set must be reduced before logical conclusions and proper inferences can be drawn.

(3) The utilization of war gaming data as a vehicle for making decisions proceeds by properly segregating the data and applying tests to eventually determine ranks for each unit across missions as the treatment varies. This reduces the overwhelming amount of initial data into manageable

sets. These sets can be used by the analyst to determine unit effectiveness (or weapons effectiveness) and, eventually, the total force effectiveness as treatments are varied. This process can be thought of as an effort to make the decision maker's work more tenable.

(4) Figure 2-11 presents a logical flow diagram for statistical analysis.

12. APPLICATION OF ANALYSIS METHODOLOGY. The analysis methodology may be best explained with reference to Figure 2-11. As indicated in Figure 2-11 there are three phases.

a. Phase I:

(1) Collect game output data and array results as shown in Figure 2-9.

(2) Begin with unit 1, activity type 1, and effectiveness indicator 1; i.e., ijk = 111 in the block data scheme.

(3) For this case, construct an array of engagement data versus treatment type.

(4) For the array constructed, assign ranks and apply the Kruskal-Wallis one-way ANOVA followed by the Mann Whitney U-Test. This will give a set of ranks for unit 1, activity type 1, and effectiveness indicator type 1, which should be recorded. At this point an inference can be made as to how treatments influenced engagement outcome. This will aid the overall analysis.

(5) Proceed to effectiveness indicator type 2 for unit type 1 and activity type 1. For the resulting data set (ijk = 121) apply the logic presented in Steps 3 and 4. Reiterate the procedure for each effectiveness indicator.

(6) After exhausting the set of effectiveness indicators for unit type 1 and activity type 1, change unit type and consider the data in the data block, ijk = 211, following Steps 3 and 4. Reiterate this procedure until all effectiveness indicator data for unit type 2 and activity type 1 have been exhausted.

(7) Using steps equivalent to Step 6, treat all data for unit types 1 through 5 with a fixed activity type.

(8) Next, move down one row and begin with the data set ijk = 112. Using Steps 4, 5, and 6, treat the remaining data sets (e.g., 122 and 132).

(9) Next, change unit type and consider data sets ijk = 212, 222, and 232. Reiterate for data sets 312, 322, and 332; 412, 422, and 432; and 512, 522, and 532.

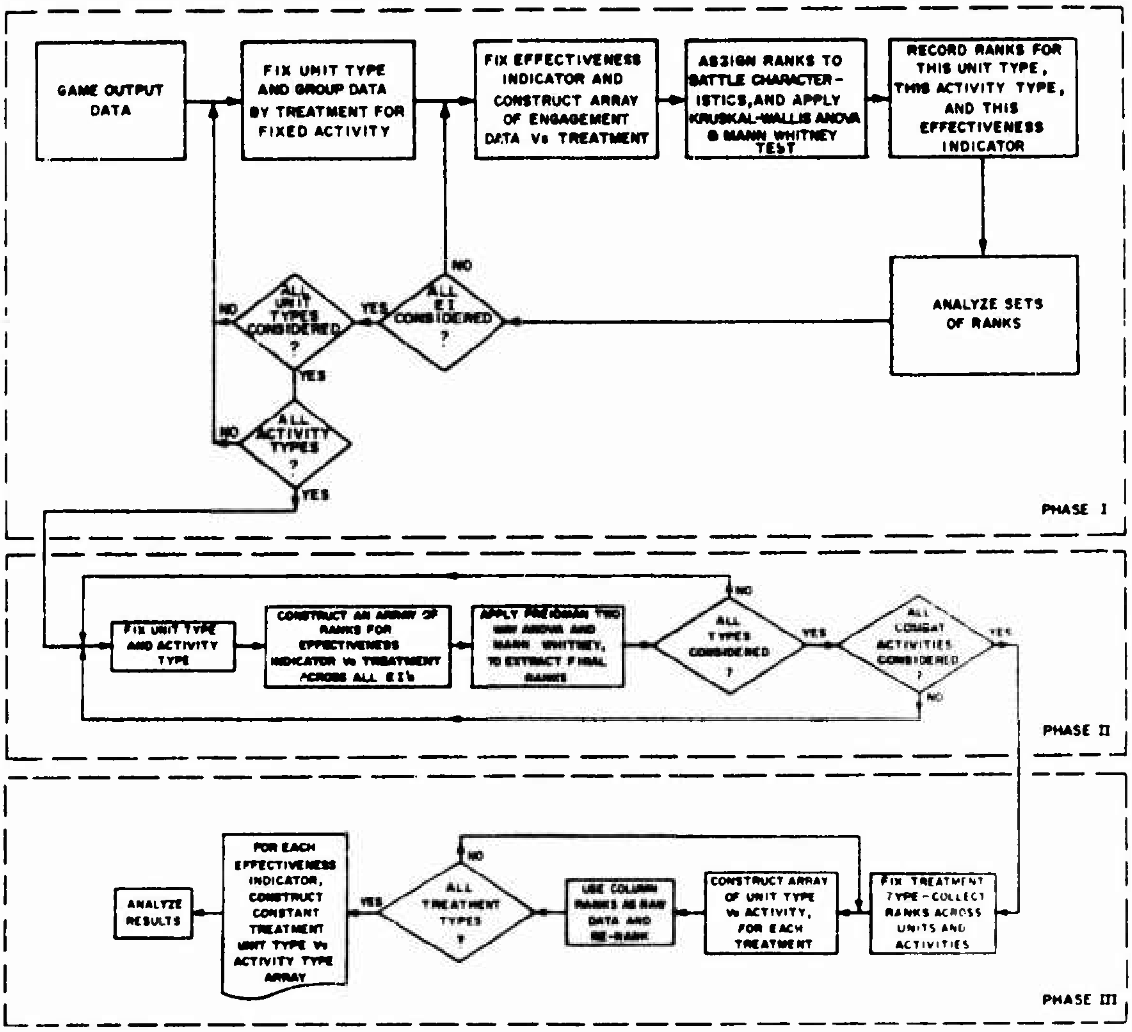


Figure 2-11. Logic Flow for Statistical Analyses

(10) Next, move down one row and continue the process with data sets ijk = 113, 123, and 133; 213, 223, and 233; and 313, 323, and 333.

(11) This process is continued until each unit type has been analyzed across all effectiveness indicators for each activity type.

b. Phase II:

(12) We now collect the ranks for unit 1 and activity 1. The array shown in Figure 2-12 is then subjected to a two-way ANOVA (Friedman and U-Test) to acquire final ranks for unit 1 and all treatment types.

	EI_1	EI_2	EI_3	EI_4	EI_5
Treatment 1	1	4	1	2	1
Treatment 2	2	1	2	3	2
Treatment 3	3	3	4	4	3
Treatment 4	4	2	3	1	4

Note: EI = Effectiveness Indicator

Figure 2-12. Calculated Ranks by Activity Type and Indicator Type for Fixed Combat Activity and One Unit

(13) Repeat Step 12 for each unit type and activity type; i.e., proceed across the rows in Figure 2-9. Then, change activity type and repeat Steps 12 and 13. Continue until all entries in each data block have been reduced to a single set of ranks. The resultant array is two-dimensional and similar to that shown in Figure 2-13.

	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5
Activity 1	Ranks	Ranks	Ranks	Ranks	Ranks
Activity 2	Ranks	Ranks	Ranks	Ranks	Ranks
Activity 3	Ranks	Ranks	Ranks	Ranks	Ranks
Activity 4	Ranks	Ranks	Ranks	Ranks	Ranks

Figure 2-13. Ranks by Treatment Type

c. Phase III:

(14) Fix the treatment type and extract the rank by unit type and activity type from Figure 2-13. This results in a single entry in Figure 2-14 for each rank.

	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5
Activity 1	Rank	Rank	Rank	Rank	Rank
Activity 2	Rank	Rank	Rank	Rank	Rank
Activity 3	Rank	Rank	Rank	Rank	Rank
Activity 4	Rank	Rank	Rank	Rank	Rank

Figure 2-14. Unit Ranks by Activity Type for a Fixed Treatment Type

For example, if in Step 13, the result for four treatment types was as shown in Figure 2-15,

	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5	
Activity 1	1 3 4 2	1 2 3 4	2 4 1 3	3 1 4 2	1 2 3 4	Treatment Type 1 Treatment Type 2 Treatment Type 3 Treatment Type 4
Activity 2	2 1 3 4	4 3 1 2	2 1 3 4	1 2 3 4	1 2 3 4	
Activity 3	1 2 3 4	1 2 3 4	1 2 3 4	3 1 2 4	1 3 2 4	
Activity 4	1 2 3 4	2 1 3 4	2 4 1 3	2 3 1 4	1 2 3 4	

Figure 2-15. Example of Ranks by Treatment Type

the new array for Step 14 for treatment 1 would be as shown in Figure 2-16.

	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5
Activity 1	1	1	2	3	1
Activity 2	2	4	2	1	1
Activity 3	1	1	1	3	1
Activity 4	1	2	2	2	1

Figure 2-16. Example of Unit Ranks by Activity Type for a Fixed Treatment Type

(15) The set of ranks acquired in Step 14 is used as raw data. Each column entry is reranked from 1 through 4.

(16) Return to Step 13, select a different treatment type, and proceed until all treatment types are exhausted.

(17) The final set of ranks is displayed as indicated in Figure 2-17.

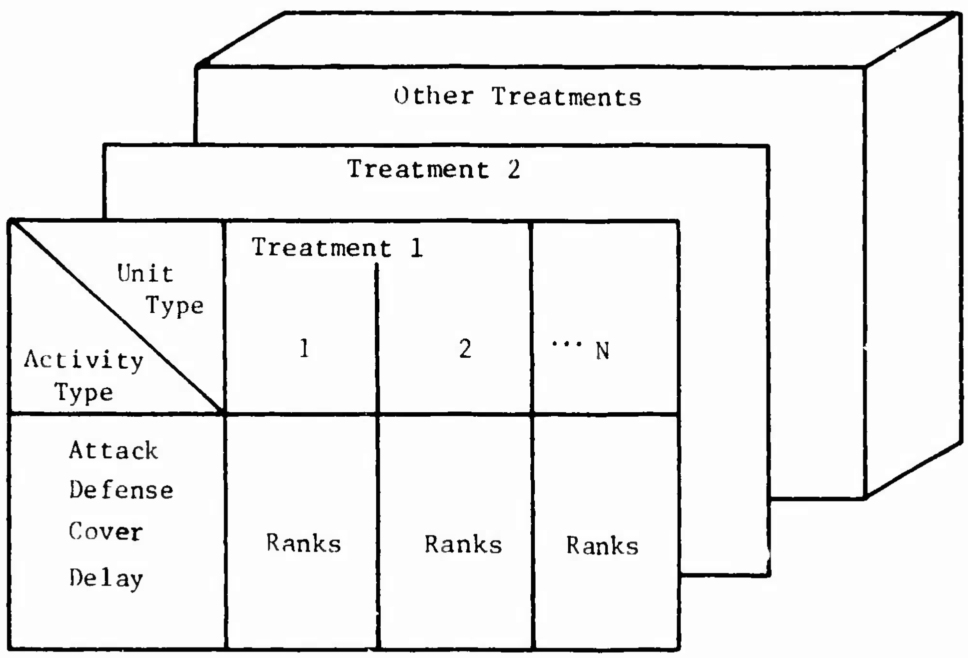


Figure 2-17. Final Arrays by Treatment Type

d. Results. This final configuration indicates by treatment how well each unit performed; thus, unit strengths and weaknesses and, therefore, force structure strengths and weaknesses may be related directly to the treatment. This allows the analyst to make decisions that have more than a purely judgmental basis and presents supporting information for analysis of the primary measure of effectiveness, mission accomplishment. Section IV presents an example of the application of the evaluation methodology to data generated through a simulation process.

Section IV. EVALUATION METHODOLOGY EXAMPLE

13. INTRODUCTION. This section presents an example of evaluation methodology procedures using performance data generated by combat simulation. The analysis is for example purposes only, and the results cannot be considered definitive for reasons pointed out in the discussion. The example is intended primarily to provide an understanding of the kinds of inferences to be drawn from statistical analysis and the synthesis of these results with subjective analysis to produce an evaluation of force effectiveness.

14. EXAMPLE. The following example presentation is keyed to the evaluation methodology logic flow, Figure 2-1. Where the problem or the example data were insufficient to support specific steps of the logic flow, their application is described relative to a more extensive analysis problem.

a. Input - Phase I. The results of a 19-hour period of continuous combat simulation generated by the DIVTAG II war game model were selected for the example analysis. The period was originally prepared and run for the purpose of testing the validity of DSL orders issued by gamers for an extended battle time period. The scenario was provided by CSC-CDRO analysts, and the data were selected from the FARMWAG data base. The 3d Armored Division, as organized and equipped in FARMWAG, was simulated in the 19-hour period against a Red threat as postulated in the FARMWAG game.

b. Game Preparation - Phase II (Steps 1 through 6):

(1) Analyze Objectives - Step 1. Although the purpose of the example evaluation is to illustrate procedures, the objective of the example analysis can be stated as the determination of the degree to which the Blue force could successfully conduct a position defense against an attacking Red force in the Fulda Gap area of Germany.

(2) Select Secondary MOEs - Step 2. Analysis and evaluation of the firepower function was selected as the best vehicle for illustrating the methodology clearly and concisely; therefore, the secondary MOE, personnel casualties and key equipment losses per unit of time as a percent of the total force, was used.

(3) Select Effectiveness Indicators - Step 3. The effectiveness indicators supporting the secondary MOE are based on personnel casualties and tank and APC losses for both Red and Blue for each battle period:

- Percent of Red personnel killed = R_c
- Percent of Red tanks killed = R_t
- Percent of Red APCs killed = R_{APC}

- Percent of Blue personnel killed = B_c
- Percent of Blue tanks killed = B_t
- Percent of Blue APCs killed = B_{APC}

These effectiveness indicators were then divided by the battle time for each engagement to extract:

- Rate of total Red personnel killed - Rate of R_c
- Rate of total Red tanks killed - Rate of R_t
- Rate of total Red APCs killed - Rate of R_{APC}
- Rate of total Blue personnel killed - Rate of B_c
- Rate of total Blue tanks killed - Rate of B_t
- Rate of Blue APCs killed - Rate of B_{APC}

absolute exchange ratios and rates:

- $\frac{\text{Total Red personnel killed}}{\text{Total Blue personnel killed}}$ - $\frac{\text{Red rate}}{\text{Blue rate}}$
- $\frac{\text{Total Red tanks killed}}{\text{Total Blue tanks killed}}$ - $\frac{\text{Red tank rate}}{\text{Blue tank rate}}$
- $\frac{\text{Total Red APCs killed}}{\text{Total Blue APCs killed}}$ - $\frac{\text{Red APC rate}}{\text{Blue APC rate}}$

and relative exchange ratios and rates:

- R_c/B_c
- R_t/B_t
- R_{APC}/B_{APC}
- R_c rate/B_c rate
- R_t rate/B_t rate
- R_{APC} rate/B_{APC} rate

(4) Define Performance Data Requirements - Step 4. For the restricted example, performance data requirements were limited to Blue and Red casualty and loss data by type. The data required are presented below:

- Red personnel casualties
- Red APC casualties
- Red tank casualties
- Blue personnel casualties
- Blue tank casualties
- Blue APC casualties
- Battle time
- Initial Red/Blue strengths

Section V of this chapter discusses the definition of performance data requirements, the organization of available output data, and the selection of performance data for more extensive evaluation problems.

(5) Develop Game Plan - Step 5. Since the 19-hour period subjected to analysis was run for other purposes, the development of a game plan was unnecessary. Normally, a game plan would be prepared to support the specific force evaluation objective and would consider such areas of game management as allocation of resources, time scheduling, and game record requirements.

(6) Load Game Input Data - Step 6. The data used for the example period were already loaded in the FARMWAG data base and were modified only slightly to meet the original test requirements.

c. Data Evaluation - Phase III (Steps 7 through 18):

(1) Select Force Mission - Step 7. The missions in which the test forces are simulated is a function of the scenario developed for the game. The evaluation methodology logic flow provides for the evaluation phase to be entered by selection of a mission for simulation, with Steps 7 through 15 being repeated for different missions until sufficient data are obtained to substantiate a determination of overall force effectiveness. The mission of the Blue force simulated in the 19-hour test period was to defend at prepared positions.

(2) Exercise Model - Step 8. The DIVTAG II war game model was exercised to produce the data analyzed in the example evaluation. Many errors in the mathematical formulation of DIVTAG II equations describing particular

physical effects have been noted (IMPWAG study, Appendix C); therefore, the analysis results are not definitive, but the output can be used as a vehicle for illustrative purposes.

(3) Selec a Combat Activity - Step 9:

(a) As discussed in Sections II and III, the data sets subjected to analysis consist of performance data produced by simulation of the distinct combat activities engaged in by subordinate elements of the total force. After the model is exercised to simulate force performance in a single mission assignment, Step 9 of the methodology calls for selection of a single combat activity to constitute a data set for analysis. Steps 9 through 12 are repeated for a different combat activity until all activities simulated during the mission are analyzed. The combat activity data set is subset on the basis of battle variables, such as terrain, visibility, and combat support.

(b) The example period provided only one combat activity for the Blue force, a defense during which Blue held its position or withdrew only slightly and reinforced when in trouble. During the period Red advanced and reinforced to maintain or enhance the advancement. The combat activity simulated during the period included four battles as depicted in Figure 2-18. The holding attack in the north was Battle ALFA, the secondary attack was Battle BRAVO, the holding attack in the center through Fulda was Battle CHARLIE, and the main attack was Battle DELTA.

(c) Battle ALFA was fought as a holding action by Red because of the difficult terrain together with the fact that a successful attack in that area would have no place to go with respect to the overall Red strategy. Accordingly, it was the mission of Red to maintain that amount of pressure necessary to contain Blue and even to cause Blue to reinforce if this were feasible with the relatively limited resources available to the attacker. For these purposes, the Red attacker was allotted two motorized rifle battalions (MRB) for the initial assault; however, one tank battalion from the reserve was alerted to reinforce should losses exceed a threshold. The threshold was set at the equivalent of three rifle companies. Using the controlled pressure tactics prescribed, Red fought more than 10 hours before reinforcement was required. With the reinforcement by the tank battalion, Red stepped up the pressure. Some 6 hours later Blue had been attrited to the point that he had to give up ground or reinforce. Because the northern Blue brigade was in an area defense, and since adjacent units were containing the enemy, he chose to reinforce.

(d) Battle BRAVO was conceived as a complicated battle, with Blue suffering sufficient losses initially to permit a Red penetration, which would have been countered by a Blue counterattack with a tank heavy battalion, to be countered, in turn, by the commitment of two Red tank battalions from the reserve. Loss rates in the BLF model did not go as predicted; accordingly, there were no reinforcements, and the battle continued unrealistically.

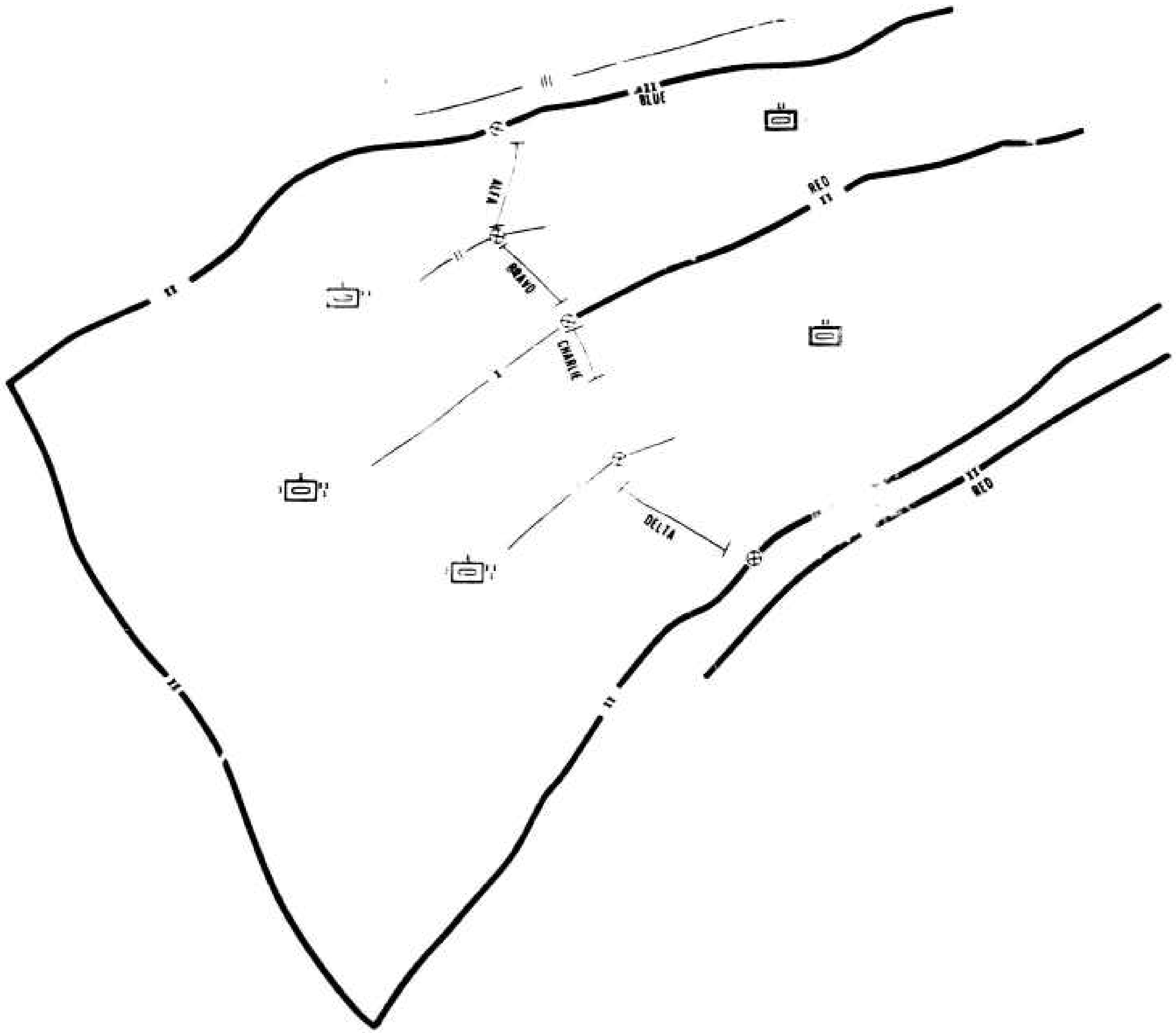


Figure 2-18. Battles of Example Game Period

(e) Battle CHARLIE, a holding attack as conceived by Red, nevertheless had adequate power to constitute a serious threat to the Blue defenses had the Blue attack helicopters (Cobras) not taken a heavy toll of Red's armored vehicles. The devastating action of these Cobras resulted in Red breaking off the engagement.

(f) Battle DELTA was designed as the main attack; accordingly, it was heavily weighted with tank strength. Blue Cobra attacks effectively reduced this tank strength; nevertheless, Red was able to force Blue into his first delaying position while Red penetrated the Blue area to a depth of some 7 kilometers.

(g) In summary, the battles constituting the 19-hour test period are as follows:

1. ALFA:

a. Phase I (10 hours 26 minutes). Blue mechanized infantry battalion task force attacked by two Red motorized rifle battalions.

b. Phase II (5 hours 54 minutes). Red reinforced with a tank battalion to carry on the attack.

c. Phase III (2 hours 40 minutes). Blue reinforced by a tank battalion.

2. BRAVO (19 hours). Blue mechanized infantry battalion task force defended against an attack by one Red tank battalion and one Red motorized rifle battalion.

3. CHARLIE:

a. Phase I (1 hour 44 minutes). Blue armored battalion task force supported by two attack helicopter troops defended against two Red tank battalions.

b. Phase II (3 hours 56 minutes). Red, reinforced by one tank battalion, fights Blue less Cobra.

c. Phase III (13 hours 19 minutes). Red discontinues attack but sustains losses due to Blue artillery.

4. DELTA:

a. Phase I (2 hours 52 minutes). Blue mechanized infantry battalion task force supported by one attack helicopter troop defends against Red attack by three motorized rifle battalions and one tank battalion.

b. Phase II (1 hour 50 minutes). Red loses one tank battalion but is reinforced by three new tank battalions and forces Blue to withdraw. Blue no longer has attack helicopter support.

c. Phase III (1 hour 11 minutes). Blue defends at first phase line after withdrawal.

d. Phase IV (12 hours 54 minutes). Blue reinforces at first phase line with one armored battalion task force and holds position.

(h) A subjective analysis of the results of the 19-hour period indicates that both Blue and Red must be considered as successful in mission accomplishment. The rationale for this conclusion is as follows:

1. In the northern Blue brigade the mission was area defense. At the end of the period Blue was in his original positions. Red's mission in Battle [illegible] was containment. That Red was successful is evidenced by the fact that [illegible] committed his reserve battalion in a reinforcing role.

2. In the southern Blue brigade the Battle DELTA mission was delay. At the end of the period the force containing the main Red attack was defending at his first delay position, even though he was reinforced at that position by the Blue brigade reserve. The other delaying battalion was in his original defensive position.

3. This relative success by Blue was very costly to one of his predominant weapon systems, the Cobra. Even though he killed Red tanks at a rate of 6 to 1 Cobra lost, he nevertheless lost 81 percent of this asset.

(4) Extract Performance Data Set - Step 10:

(a) With the current version of the DIVTAG II model output, the data extraction effort was of primary importance since it required the most time. The effort involved in extracting performance data from the printouts of the test battles conducted was considerable. The reports used were:

- Assessment Report
- Battle Engagement Report
- Movement Report
- Unit Status File Dump
- DSL Printout

(b) As each battle was traced through its course, the points in that battle were noted when either side took an action that had the effect of changing the battle conditions. These points were deduced from analyses of DSL (which said what the player wanted to do under various conditions), the Movement Report (which showed who moved, when, to where, and how often he was fired on), and the Battle Engagement Report (which showed the location of each engagement and the time it occurred as well as the results). This process entailed much maneuvering as the various data were studied. For instance, it was noted that Battle ALFA had a variety of time intervals between engagements. The interval was greater as time went by and losses mounted, but midway through the period the time interval dropped from 61 minutes to 16 minutes. A study of all the output revealed that just prior to this point, Red was reinforced with a tank battalion; thus, a new set of battle conditions existed. For statistical analysis purposes, summaries of performance data up to that point had to be laboriously tabulated from the Battle Engagement Report, the Movement Report, and the Assessment Report to arrive at the status of units and their losses from various causes. Later in the same battle, Blue reinforced with a tank battalion; and, again, a new set of battle data up to that point had to be extracted manually.

(c) With the aid of considerable experience in DIVTAG II output, and by a process of trial and error, the following subsets of data were extracted, from which the statistical analysis could be performed:

- Battle ALFA - 3 subsets
- Battle BRAVO - 1 subset
- Battle CHARLIE - 2 subsets (plus a period of losses to artillery only)
- Battle DELTA - 2 subsets
- Battle DELTA 1 - 2 subsets

(d) In this test case, the conditions under which a summary of performance was prepared and a new battle was considered to have begun were:

- Either Red or Blue committed an additional unit to the engagement
- Either Red or Blue withdrew or discontinued the attack

(e) The data set extracted was small; however, since the problem of extraction of data from the current DIVTAG II output format was so time consuming, only a limited number of effectiveness indicators was used. The data set is presented in Step 4 and was used to calculate the designated effectiveness indicators presented in Step 3.

(f) The effectiveness indicator data were then arrayed. Since Red had only one combat activity--attack--and Blue had only one combat activity--defend, the analysis across combat activities could not be made. Also, because DIVTAC II is not sufficiently versatile, it was impossible for the analysis to consider variations in weather, terrain, light level, mobility, and other background conditions which contribute to the effectiveness of a force. Furthermore, for this test, only a few units participated; therefore, rather than testing background conditions, engagement data were collected by unit type with no variation in treatment. Three unit structures were used on the Blue side:

. Mechanized

. Armor

. Mechanized and armor

and four unit structures were used on the Red side:

. Motorized

. Armor

. Motorized and armor

. Motorized, armor, and air defense.

The data set, therefore, was arrayed by effectiveness indicator and unit type as shown in Figure 2-19.

(5) Conduct Statistical Analysis and Interpret Analysis Results - Steps 11 and 12. After the output data were organized into proper subsets, the statistical analysis was initiated. The following description of the statistical analysis is keyed to the statistical analysis diagram, Figure 2-11 of Section III, and the accompanying discussion; however, since the example period included only one combat activity for both Red and Blue, the full analysis procedures could not be performed. Only the first four steps of Phase I of the three-phase procedure (Figure 2-11) were carried through. Other steps were not used because of a lack of data. The data allowed only the application of the Kruskal-Wallis one-way analysis of variance. The lack of data prevented the application of the Mann-Whitney U-Test and the initiation of pairwise testing. The following paragraphs describe stepwise the statistical analysis procedures applied under Phase I of the analytical methodology (Section III) and present an interpretation of the results for example purposes.

(a) Array Game Output Data. As described for Step 10, the game output data were arrayed by unit type versus engagement outcome. Three blocks of data were used where each block pertains to a fixed unit type. In this respect the 1j1, 2j1, and 3j1 blocks were of interest. In each case

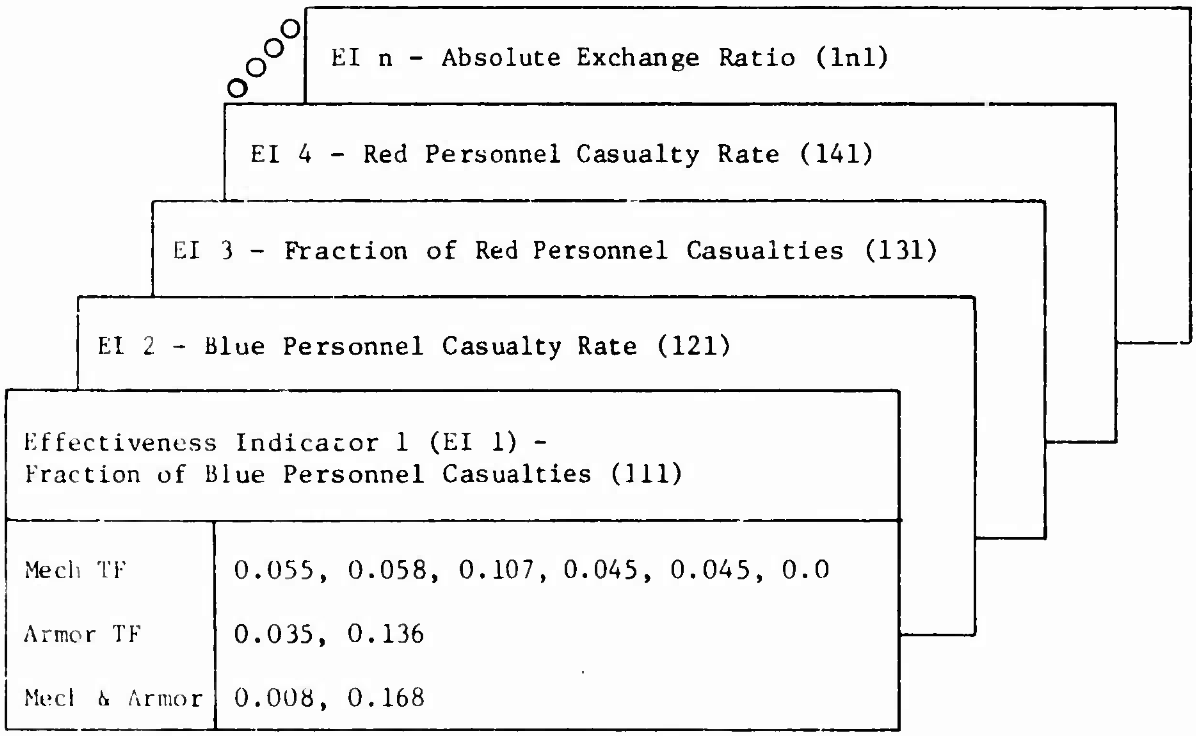


Figure 2-19. Data Array for Blue Defense

j was allowed to have all values from 1 to as many indicators as were desired. There are two separate and distinct sets of blocks of this kind. One pertains to engagement results using Blue units, and the other relates to Red units. For the Red units four blocks were used.

(b) Begin with Unit 1, Activity 1, and Effectiveness Indicator 1 (ijk=111). This step is done for both Red and Blue. This produces only one row of engagement data; therefore, it is necessary that this row's counterpart for all units be considered. When j=1 we are interested in personnel. Furthermore, for this example, only the Blue mix is of current interest; therefore, casting the existing data into a set of three rows by unit type allows an analysis across units.

(c) Construct an Array of Engagement Data versus Treatment Type. For the case selected, fractional number of personnel lost, we have (Figure 2-20):

Battle	Units	Fractional Number of Personnel Lost
ALFA 1	Mech TF	0.055
ALFA 2	Mech TF	0.058
BRAVO	Mech TF	0.107
DELTA 1	Mech TF	0.045
DELTA 2	Mech TF	0.045
DELTA 3	Mech TF	0.0
CHARLIE 1	Armor TF	0.035
CHARLIE 2	Armor TF	0.136
ALFA 3	Mech & Armor	0.008
DELTA 4	Mech & Armor	0.168

Figure 2-20. Engagement Data versus Treatment Type

An analysis based on a fixed unit type is impossible, as it presents only a single row of data; therefore, all engagement results for personnel are considered. Thus, we are concerned with the 111, 211, and 311 data sets and treat them as a single set. This allows an analysis across unit types.

(d) Assign Ranks and Apply the Kruskal-Wallis ANOVA. (Figure 2-21.) The null hypothesis:

H_0: There is no difference [illegible] fractional personnel casualties as the Blue force mix is [illegible]ered,

will be tested against its alternative:

H_1: There is a difference in fractional personnel casualties as the Blue force mix is altered.

Units	Ranks Assigned	Rank Sums
Mech TF	6 7 8 5 4 1	$n_1 = 6$ $R_1 = 31$
Armor TF	3 9	$n_2 = 2$ $R_2 = 12$
Mech & Armor	2 10	$n_3 = 2$ $R_3 = 12$

Figure 2-21. Assignment of Ranks for Kruskal-Wallis ANOVA

Calculation then proceeds as follows:

$$H = \frac{12}{N(N+1)} \sum_i \frac{R_i^2}{n_i} - 3(N+1)$$

$$N = n_1 + n_2 + n_3 = 10$$

So:

$$H = \frac{12}{10(11)} \left\{ \frac{(31)^2}{6} + \frac{(12)^2}{2} + \frac{(12)^2}{2} \right\} -3(11)$$

$$H = \frac{12}{10(11)} \left\{ 304 \right\} - 33 = 33.2 - 33 = 0.2$$

From Seigel (reference 1), Table C, page 249, it may be seen that this null hypothesis must be accepted. The inference here is that if there is a difference that should be noted, the model output is insensitive to this difference. Two procedures are possible at this point. It is possible to establish a significance level and accept or reject the null hypothesis. Conversely, it is possible to assume that the null hypothesis is always rejected and use the value of H with Table C to determine the significance

level α . In this latter case (which will be used extensively in this test) the value of α represents the probability of making an error when it is assumed that the null hypothesis must be rejected. When we say "accept H_0," this should be interpreted to mean that when we assume H_0 should be rejected the calculated value of α is too large to warrant rejection at any level.

(c) Reiterate This Procedure Across All Effectiveness Indicators of Interest:

1. Using the Blue organization as a basis, the following results were extracted from the analysis:

- Percent of Red personnel killed - Reject H_0 at $\alpha = 0.001$
- Percent of Red tanks killed - Accept H_0
- Percent Red APCs killed - Accept H_0
- Percent of Blue personnel killed - Accept H_0
- Percent of Blue tanks killed - Accept H_0
- Percent of Blue APCs killed - Accept H_0

The result for Red personnel seems out of proportion when considered with the other results. An error in coding has been found in the DIVTAG II casualty assessment scheme wherein the Red targets killed were not made absolute; i.e., only conditional Red casualties were calculated in the past. This error was found in building the new ground combat model and is a possible reason for the small value of α calculated for Red personnel. To reiterate, reject at 0.001 means that when we reject the null hypothesis we have only a 0.001 chance of making an error.

2. For these end of engagement results we now consider the Red organization. The null hypothesis:

H_0: There is no difference in engagement outcome across all Red organizations tested,

and its alternative:

H_1: There is a difference in engagement outcome across all Red organizations tested,

must be tested using the Kruskal-Wallis Test. Using standard procedures the following conclusions can be drawn for these data:

- Percent of Red personnel killed - Accept H_0
- Percent of Red tanks killed - Accept H_0
- Percent of Red APC killed - Reject H_0 at $\alpha = 0.70$
- Percent of Blue personnel killed - Reject H_0 at $\alpha = 0.70$
- Percent of Blue tanks killed - Reject H_0 at $\alpha = 0.30$
- Percent of Blue APCs killed - Reject H_0 at $\alpha = 0.20$.

From these results we may infer that if end of engagement data are used, there are more than seven chances in ten that the difference in Blue tank and APC losses as mirrored in the data array for Red, was caused by a difference in individual Red unit performance. Other inferences might be possible from further analysis of the data.

(f) Pairwise Testing. The next logical step would be to pairwise test the data to determine which Blue/Red unit was most effective. Unfortunately, there were not enough data to perform this test, as the Mann-Whitney U-test requires at least three entries in one row of the two sets to be tested. For this reason, three battle periods are recommended for future tests.

(g) Other Analysis Results:

1. The end of engagement results are interesting but do not present the total picture of what occurred; therefore, the same data were used, along with the engagement length, to calculate kill rates and loss rates. The data for Red use the null hypothesis:

H_0: There is no difference in rates due to organization or weapon mix,

and its alternative:

H_1: There is a difference in rates due to organization or weapons mix.

Results are as follows:

- Red personnel - Reject H_0 at $\alpha = 0.02$
- Red tanks - Reject H_0 at $\alpha = 0.10$
- Red APCs - Reject H_0 at $\alpha = 0.50$

- Blue personnel - Reject H_0 at α = 0.02
- Blue tanks - Reject H_0 at α = 0.02
- Blue APCs - Reject H_0 at α = 0.15

Even though an error in the calculation of Red casualties was noted, the fact that in all cases the significance level at which we are allowed to reject H_0 is small, requires that we infer that loss rate and casualty generation rate are in some way different as the Red unit type is changed. Because of the lack of data, this inference cannot be pursued to the point of determining which unit had the greatest impact.

2. Similar analysis of rates using Blue organization structure results in the following findings:

- Fraction Red personnel killed/unit time - Reject H_0 at α = 0.4
- Fraction Red tanks killed/unit time - Accept H_0
- Fraction Red APCs killed/unit time - Reject H_0 at α = 0.05
- Fraction Blue personnel lost/unit time - Reject H_0 at α = 0.4
- Fraction Blue tanks lost/unit time - Accept H_0
- Fraction Blue APCs lost/unit time - Accept H_0

The only significant value of α pertains to the rate of kill of Red APCs. This result is evidently due to the error in Red casualty calculations. If not, it implies that Red committed significantly more APCs that eventually became targets than it did either personnel or tanks.

3. Oftentimes the analyst/gamer is interested in the price he must pay in terms of materiel to acquire a particular objective. This is calculated using exchange ratios. There are two types of exchange ratios. The absolute exchange ratio is the ratio of Red to Blue casualties, whereas the relative exchange ratio is the ratio of fraction of Red that become casualties to fraction of Blue that become casualties. We begin with the absolute exchange ratio. For the Red force the data yield results for the following null hypothesis:

H_0: There is no difference in absolute exchange ratio due to Red force structure,

and its alternative:

H_1: There is a difference in absolute exchange ratio due to Red force structure.

The results for the data are:

$\frac{R_c}{B_c} = \frac{\text{Red personnel killed}}{\text{Blue personnel killed}}$ - Reject H_0 at $\alpha = 0.15$

$\frac{R_T}{B_T} = \frac{\text{Red tank killed}}{\text{Blue tank killed}}$ - Accept H_0

$\frac{R_{APC}}{B_{APC}} = \frac{\text{Red APC killed}}{\text{Blue APC killed}}$ - Reject H_0 at $\alpha = 0.015$

One should recall that these results are suspect because Red casualties were calculated incorrectly; however, they do support the methodology in that if there had been no error and these results had accrued, we would infer that for two of the three statistics there is an indication that performance data were influenced by organization structure.

4. Use of the relative exchange ratio tests the tradeoffs made by a force. In this case we are testing the hypothesis:

H_0: There is no difference in relative exchange ratio as the Blue organization changes,

versus the alternative:

H_1: There is a difference in relative exchange ratio as the Blue organization changes.

As may be seen, the conclusions are:

$\frac{\text{Percent } R_c}{\text{Percent } B_c}$ - Reject H_0 at $\alpha = 0.25$

$\frac{\text{Percent } R_T}{\text{Percent } B_T}$ - Accept H_0

$\frac{\text{Percent } R_{APC}}{\text{Percent } B_{APC}}$ - Accept H_0

Since α is large in each case the inference to be made is that there is no significant difference in the Blue performance as the Blue unit type is varied. The error in the Red casualty calculation could have caused this set of results. For Red the results are:

$$\frac{\text{Percent } R_c}{\text{Percent } B_c} \quad - \text{ Reject } H_0 \text{ at } \alpha = 0.07$$

$$\frac{\text{Percent } R_T}{\text{Percent } B_T} \quad - \text{ Reject } H_0 \text{ at } \alpha = 0.15$$

$$\frac{\text{Percent } R_{APC}}{\text{Percent } B_{APC}} \quad - \text{ Reject } H_0 \text{ at } \alpha = 0.25$$

Here it is possible to reject the null hypothesis with little probability of making an error when considering personnel; however, as tanks and APC's are considered, the inference that the results accrue as the Red unit structure is changed has a successively larger probability of being in error.

5. We now consider the same data as a function of time and calculate exchange ratio rates. The null hypothesis:

H_0: There is no difference in rates due to Blue force structure,

versus its alternative:

H_1: There is a difference in rates due to Blue force structure,

was tested. The results are:

$$\frac{R_c}{B_c} \quad - \text{ Reject } H_0 \text{ at } \alpha = 0.25$$

$$\frac{R_T}{B_T} \quad - \text{ Reject } H_0 \text{ at } \alpha = 0.25$$

$$\frac{R_{APC}}{B_{APC}} \quad - \text{ Reject } H_0 \text{ at } \alpha = 0.40$$

These results are undoubtedly affected by the inaccurate Red casualty calculation. If not, due to the large values of α, the only inference possible is that rates of exchange could become a significant statistic.

6. After these data have been analyzed, some inferences can be made. It is particularly evident that the error in coding the Red casualty calculation had impact on these findings. Furthermore, since many null hypotheses were accepted, and nearly all were accepted at the predetermined level of $\alpha = 0.10$, we must conclude that there is no significant difference in the effects of the major weapons systems assigned the Red and

Blue organizations. If this result is unsuspected at the outset, and is not judgmentally correct, we must conclude that the fire priority of the weapons systems of interest does not vary significantly from one organization to the next. This inference was verified through a check of the Red and Blue TOF data loads. Further inferences can be made. For example, there were two slightly different terrain types used in the 19-hour test period. These types were considered by the gamers, and both Red and Blue move rates were adjusted to correct for this effect.

(6) Last Combat Activity - Step 13. Step 13 in the methodology flow represents a decision point at which a determination is made that all combat activities simulated for the mission have been analyzed. If not, the flow returns to Step 9, and another activity is analyzed. If all combat activities have been analyzed, the methodology proceeds to Steps 14 and 15. Since in the example period, only one combat activity was simulated, the evaluation can move to succeeding steps.

(7) Conduct Statistical Analysis Across Combat Activities - Step 14. This step, which was not possible for the analysis example, is addressed in detail as Phases II and III of the statistical analysis methodology, Section III.

(8) Conduct Subjective Analysis Across Combat Activities - Step 15. At this point in the evaluation, military analysts would conduct a subjective analysis of force performance in all combat activities engaged in during the mission. This analysis draws upon game records, as well as the statistical analyses, reveals insights developed during game play, and tends to place purely statistical results in perspective.

(9) Sufficient Data - Step 16. Step 16 represents a decision point at which determination is made as to whether the combat activities analyzed for the single mission simulated provide a sufficient basis for a summary of overall force effectiveness. If not, the methodology recycles to Step 7, and a new force mission is simulated and its activities analyzed. When sufficient data are obtained, the force effectiveness summary is prepared, incorporating the results of any side analyses.

(10) Conduct Side Analyses - Step 17. Side analyses are various tests conducted apart from the main flow of the game to investigate areas of interest more extensively than is possible during game play. Several such analyses were conducted with data from the example period to illustrate possible applications.

(a) Sensitivity Testing for Cobra Effects:

1. One side analysis application is sensitivity testing. For example, since a portion of the Blue force used the Cobra, it is of interest to determine the effect of removal of the Cobra on the effectiveness of Red and Blue organizations. For purposes of illustration the number of kills due to the Cobra was extracted from the data. The remaining data are

not meant to be characteristic of that which would result if the Cobra were pulled from the system and the simulation rerun (although such sensitivity testing must be performed in the future). Instead, it was hoped that some further insight could be obtained as to whether total number of casualties or casualty rate would be of most importance when illustrating a difference in effectiveness.

2. The hypotheses tested were:

H_0: There is no difference in results due to Blue force structure,

and its alternative:

H_1: There is a difference in results due to Blue force structure.

The results were to accept H_0 for every statistic.

3. These data were divided by the engagement times to test the null hypothesis:

H_0: There is no difference in rates due to Blue force structure.

Results for the rates acquired were:

- Red personnel - Reject H_0 at α = 0.40
- Red tanks - Reject H_0 at α = 0.05
- Red APCs - Accept H_0
- Blue personnel - Reject H_0 at α = 0.40
- Blue tanks - Accept H_0
- Blue APCs - Accept H_0

The inference in this case is that Blue unit structure caused a significant difference in the number of tanks killed. The other values of α are too large to make a definitive statement. The tank results could have been due to the error in Red casualty assessment.

(b) Sensitivity Testing for Terrain Differences:

1. Although DIVTAG II is relatively insensitive to background conditions, the fact that the battles ALFA and BRAVO were fought over a terrain which was not as open as that for battles CHARLIE and DELTA could be a significant factor. As mentioned above, move rates and other terrain dependent activities were gamer controlled. This control was essentially all judgmental; therefore, it is of interest to see if the terrain type had any influence on the battle outcome.

2. It is possible for the analyst to check the outcome of engagements as a function of terrain by means of the Mann-Whitney U-test. Figure 2-22 contains data for battles ALFA and BRAVO (terrain type I) and battles CHARLIE and DELTA (terrain type II). In this figure, entries compared include data for both mechanized and armor units.

Terrain Type	Battle	Fractional Number of Personnel Lost
I	ALFA 1 ALFA 2 BRAVO	0.055 0.058 0.107
II	CHARLIE 1 CHARLIE 2 DELTA 1 DELTA 2	0.035 0.136 0.045 0.045

Figure 2-22. Engagement Results by Terrain Type

3. Following the procedure for the Mann-Whitney U-Test results in the set of ranks shown in Figure 2-23.

Terrain Type	Ranks Assigned	Rank Sums
I	4 5 6	$n_1 = 3$ $R_1 = 15$
II	1 7 2.5 2.5	$n_2 = 4$ $R_2 = 13$

Figure 2-23. Assignment of Ranks for Mann-Whitney U-Test

4. By definition: (for small samples)

$$U_1 = n_1 n_2 + \frac{(n_1 + 1)}{2} n_1 - R_1$$

$$U_1 = 3x4 + \frac{3x4}{2} - 15 = 3$$

$$U_2 = n_1 n_2 + \frac{(n_2+1)n_2}{2} - R_2$$

$$U_2 = 3x4 + \frac{4x5}{2} - 13 = 9$$

5. Selecting the smallest U (U_1 = 3) and testing for significance by Table J, page 271 of Seigel (reference 1), we determine α = 0.2. Carrying through with this test for all percentage losses produces values for the significance level α. In each case we test the hypothesis:

H_0: There is no difference in results due to terrain variation,

versus its alternative:

H_1: There is a difference in results due to terrain variation.

The results are as follows:

Blue personnel lost	-	Reject H_0 at α = 0.20
Blue tanks lost	-	Reject H_0 at α = 0.07
Blue APCs lost	-	Reject H_0 at α = 0.114
Red personnel lost	-	Reject H_0 at α = 0.196
Red tanks lost	-	Reject H_0 at α = 0.429
Red APCs lost	-	Reject H_0 at α = 0.196

6. This implies that there is indeed some effect due to a varying terrain. Since all data over the two terrain classes were considered, however, the difference may have been the result of the fact that different weapons mixes were used across the FEBA. For this reason only mechanized heavy units were considered across these two terrain types. The data were insufficient to consider armor heavy and a combination of the two.

7. The null hypothesis:

H_0: For mechanized heavy units, terrain differences have no influence on engagement outcome,

is to be tested against the alternative:

H_1: For mechanized heavy units, terrain differences have an influence on engagement outcome.

The results:

Blue personnel casualties	-	Reject H_0 at $\alpha = 0.05$
Blue tank casualties	-	Reject H_0 at $\alpha = 0.35$
Blue APC casualties	-	Reject H_0 at $\alpha = 0.40$
Red personnel casualties	-	Reject H_0 at $\alpha = 0.35$
Red tank casualties	-	Reject H_0 at $\alpha = 0.35$
Red APC casualties	-	Reject H_0 at $\alpha = 0.35$

indicate that there is indeed some terrain effect, but its significance is not as large as believed to be the case before the effects of varying weapons systems were removed.

(11) Prepare Overall Force Effectiveness Summary - Step 18. This step of the evaluation methodology represents a synthesis of all statistical, subjective, and side analyses to determine overall force effectiveness. The summary is oriented toward answering the analysis objectives guiding the entire evaluation. A force effectiveness summary was obviously inappropriate for the analysis example, but a summary of the test findings is presented below for illustrative purposes.

(a) A 19-hour DIVTAG II battle set was acquired for analysis in preliminary form. This analysis was restricted somewhat by the lack of data; therefore, it is recommended that at least two, and possibly three, periods of 12-hour duration be run for experimental design purposes.

(b) From the tests applied several inferences can be made. Important observations in this context are:

1. Red casualties have been calculated improperly.

2. In most instances there is very little difference in weapon system performance. This indicates that major weapons systems do not differ appreciably from one mix to the next, or that the priorities of firing for each system have been constructed in such a manner that differences in weapon systems are not apparent.

3. It appears that kill rate and casualty rate are more significant statistics than total kills and total casualties taken.

(c) The analysis encountered difficulty in extracting data from the current version of DIVTAG II. In this respect it is evident that

an efficient management information retrieval and display system (IRADS) must be constructed that will allow the extraction of all data of interest when a critical point (e.g., battle termination, reinforcement, withdrawal) is reached. Likewise, battle termination and other critical points must be tested parametrically.

Section V. ORGANIZATION OF MODEL OUTPUT DATA

15. INTRODUCTION. This section presents the complete set of arrays showing the organization of output available from the DIVWAG model. The purpose of their development is discussed, and examples of their use are provided.

16. PURPOSE. The wealth of performance data generated by the DIVWAG model in a combat simulation provides the basis for the entire evaluation methodology. On the other hand, the evaluation is structured by the MOEs and effectiveness indicators chosen to support the analysis objectives; thus, an efficient and logical method for selecting the performance data to quantify the designated MOEs and effectiveness indicators is requisite to the analysis methodology. For this reason, CSC-CDRO developed a complete set of organizational arrays for model output data keyed to the requirements of the MOE hierarchy.

17. OUTPUT DATA ARRAYS. The overall organization of model output data is shown in Figure 2-8, repeated here as Figure 2-24 for reader convenience. Subsequent figures, 2-25 through 2-44, show the organization of Blue and Red unit performance data and functional area and system performance data. Using these arrays as a guide, the analyst can select only that data output to support the MOEs and effectiveness indicators pertinent to his analysis. Other data could be retained on computer storage devices for subsequent requirements or side analyses.

18. EXAMPLE OF UTILITY:

a. As an example of the utility of the model data output arrays, consider the requirement to quantify effectiveness indicators 1 and 2 supporting the mobility function:

(1) Average speed (kilometers per hour) of all ground tactical movement by Blue (Red):

(a) Maneuver battalions.

(b) Field artillery battalions.

(c) Engineer battalions.

(2) Average speed (kilometers per hour) of all airmobile tactical movements by Blue (Red):

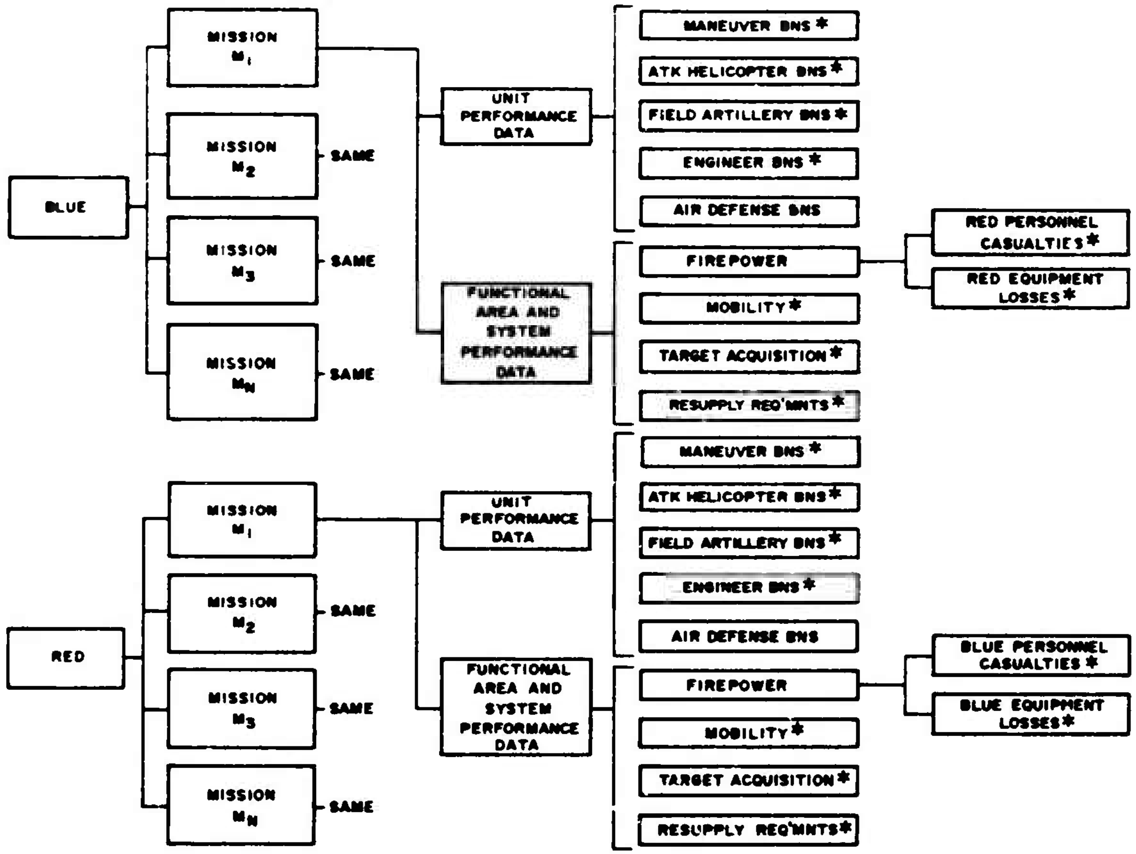
(a) Maneuver battalions.

(b) Field artillery battalions.

(c) Engineer battalions.

b. To quantify these two effectiveness indicators the analyst refers to the output data organization for Blue and Red maneuver units, field artillery units, and engineer units, Figures 2-25 through 2-30. From these arrays he selects the performance data packages for ground movement and airmobile movement (indicated on the figures by broken lines).

c. The formats in which the various performance data arrays are printed may be programmed according to the requirements of the specific force evaluation project. Effectiveness indicators may require computer calculations in addition to those producing direct game output. For example, effectiveness indicator 1 supporting the firepower function calls for the percent of all Red (Blue) personnel that became casualties as a result of Blue (Red) firepower. The output data arrays for Blue personnel casualties and Red personnel casualties, Figures 2-31 and 2-32, represent the game data sources; but the computer must perform additional calculations to quantify the performance data into specific effectiveness indicators. Figure 2-45 is an example of a performance data array for firepower; the last entry in the personnel column provides the data to quantify the effectiveness indicator. Performance data output may be arrayed in many different ways to quantify the effectiveness indicators selected to guide a specific analysis.



* - SEE FIGURES 2-25 THROUGH 2-44

Figure 2-24. Organization of Model Output Data

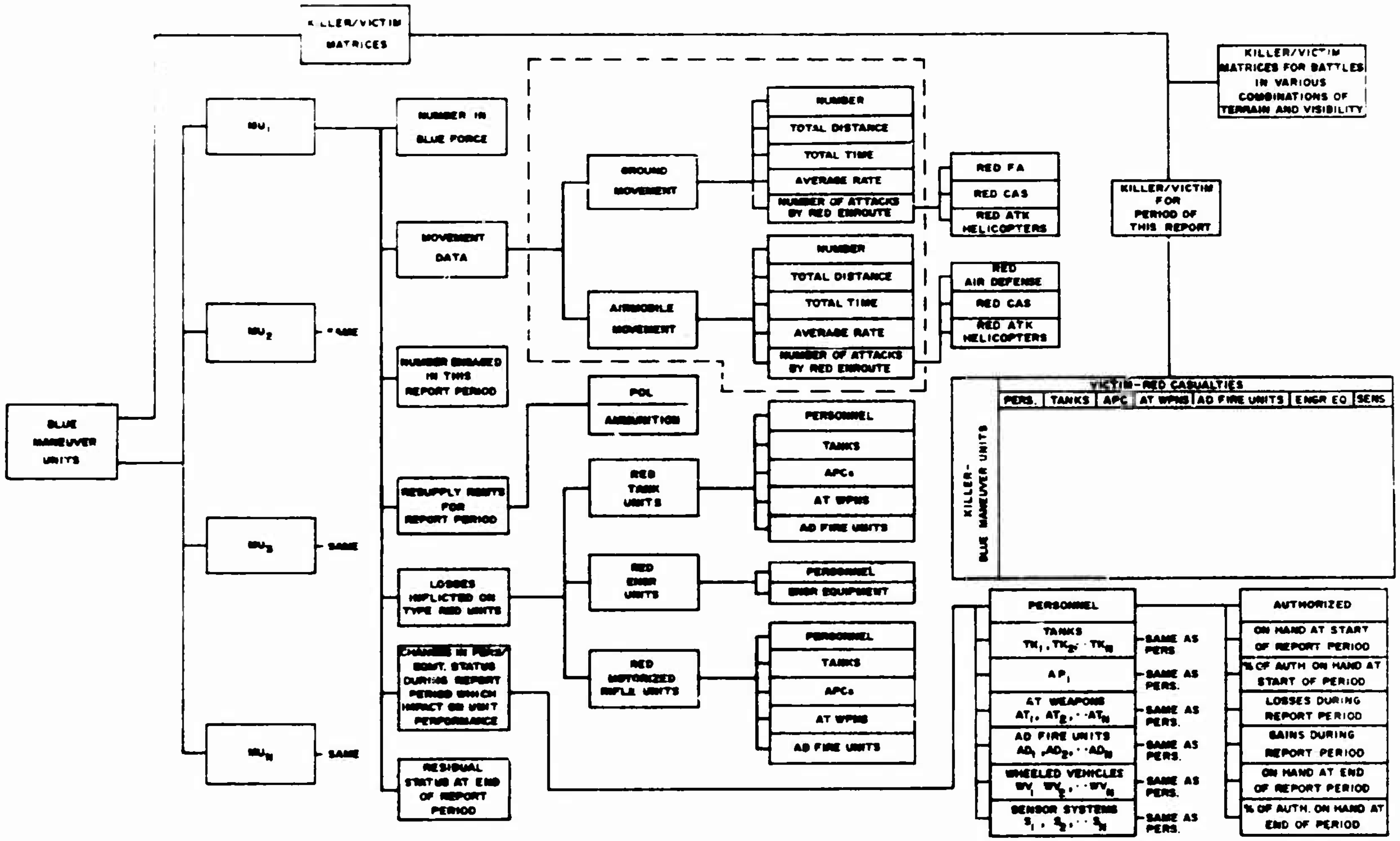


Figure 2-25. Blue Maneuver Units Output Data Array

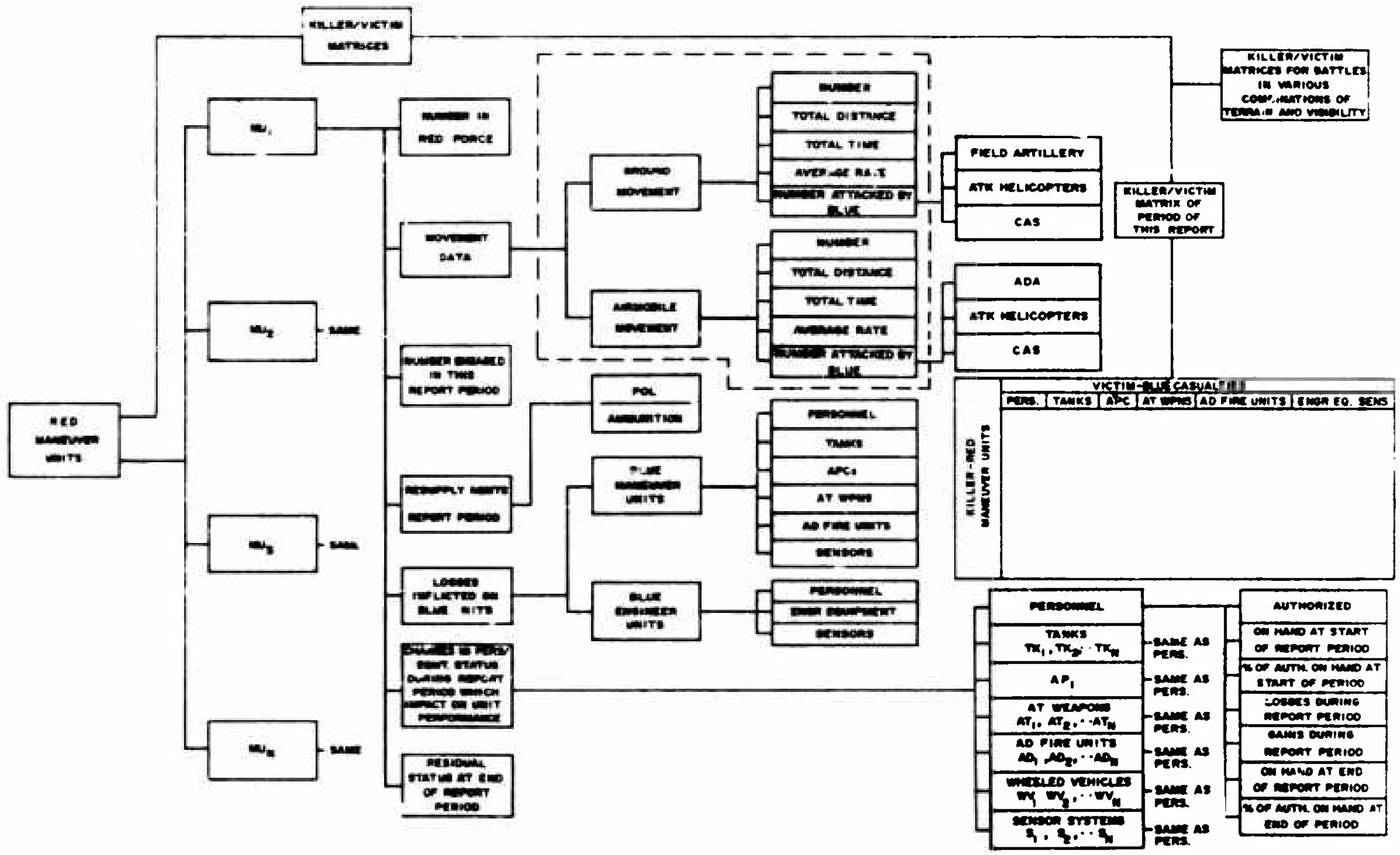


Figure 2-26. Red Maneuver Units Output Data Array

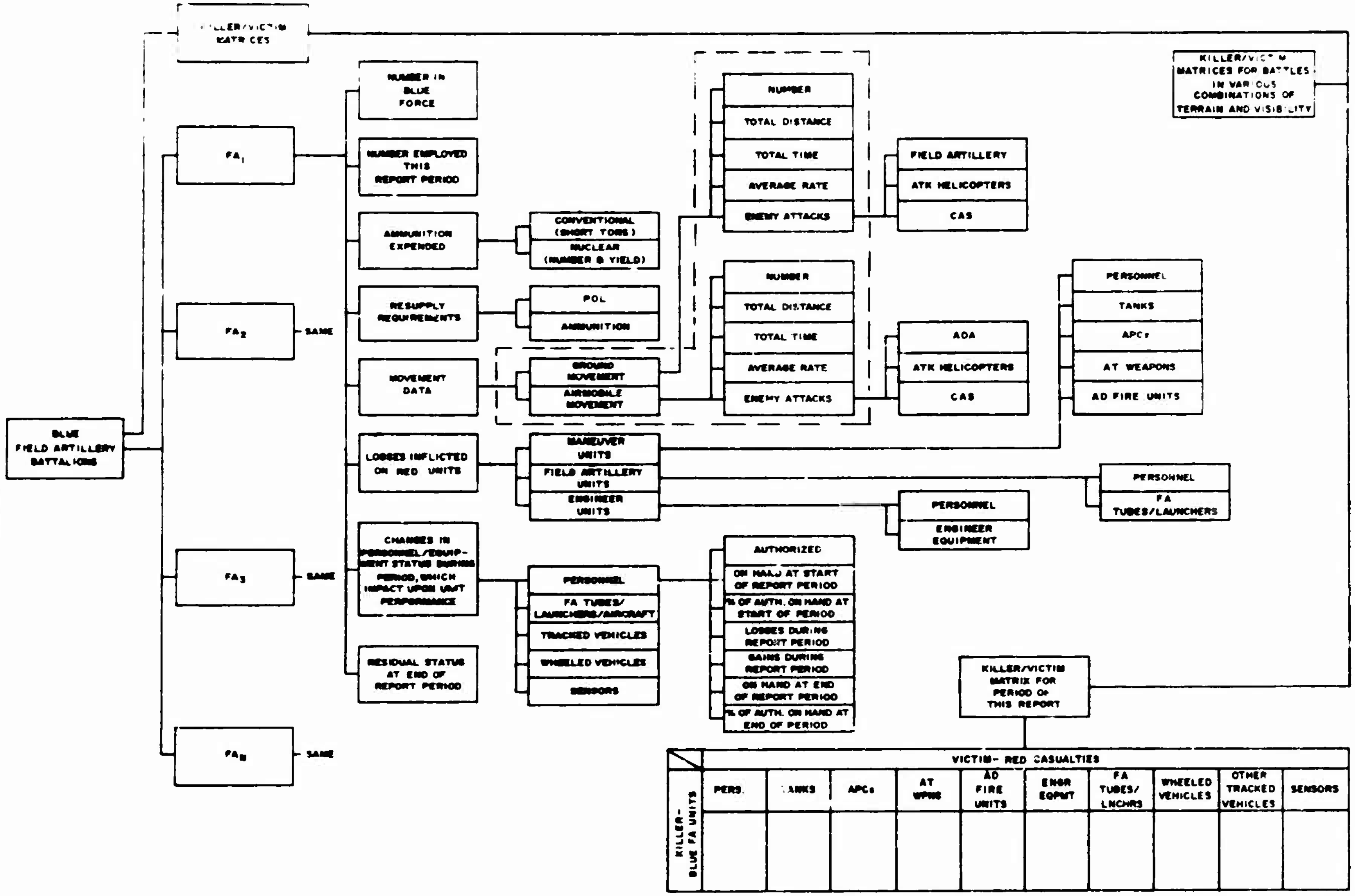


Figure 2-27. Blue Field Artillery Battalions Output Data Array

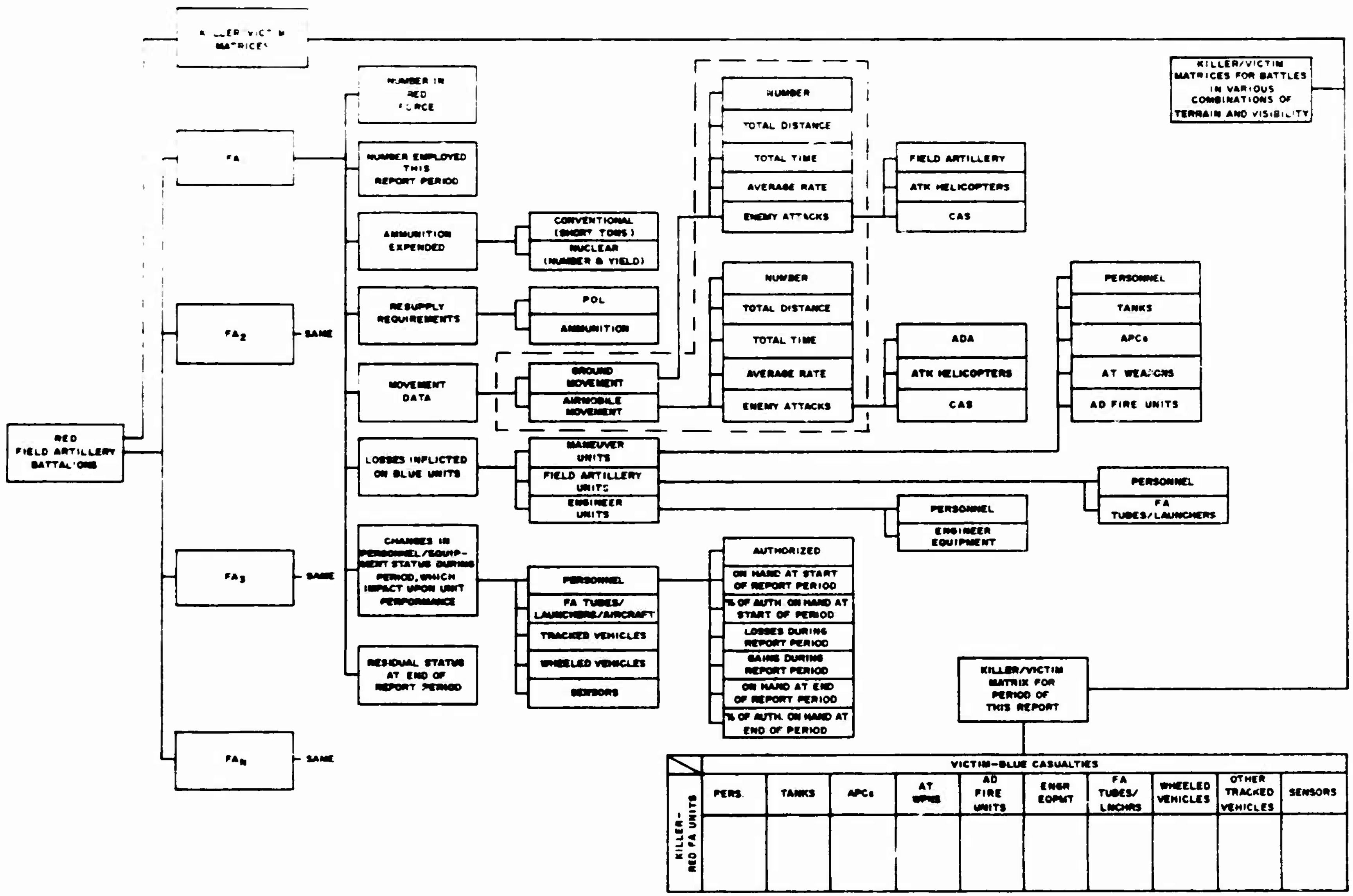


Figure 2-28. Red Field Artillery Battalions Output Data Array

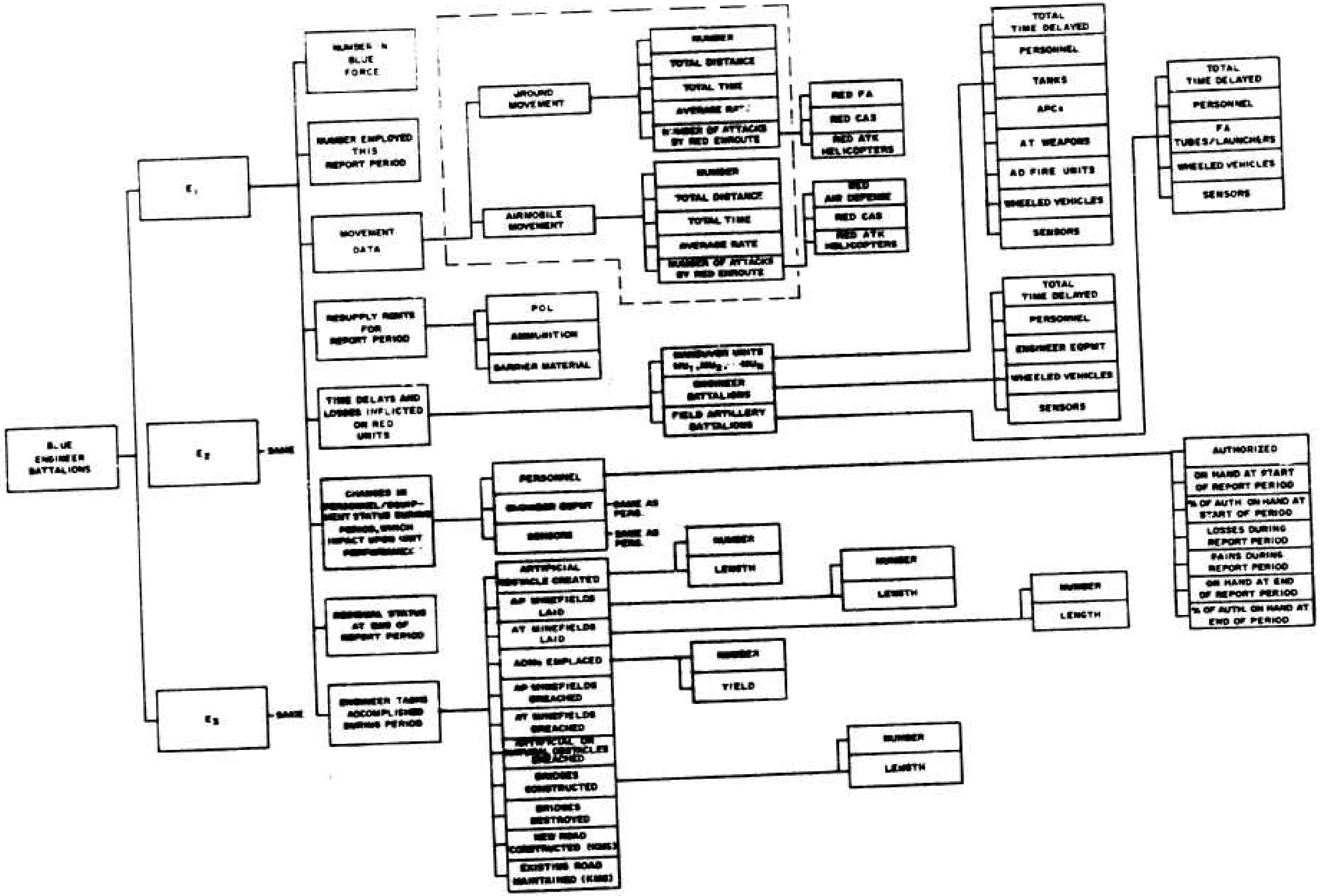


Figure 2-29. Blue Engineer Battalions Output Data Array

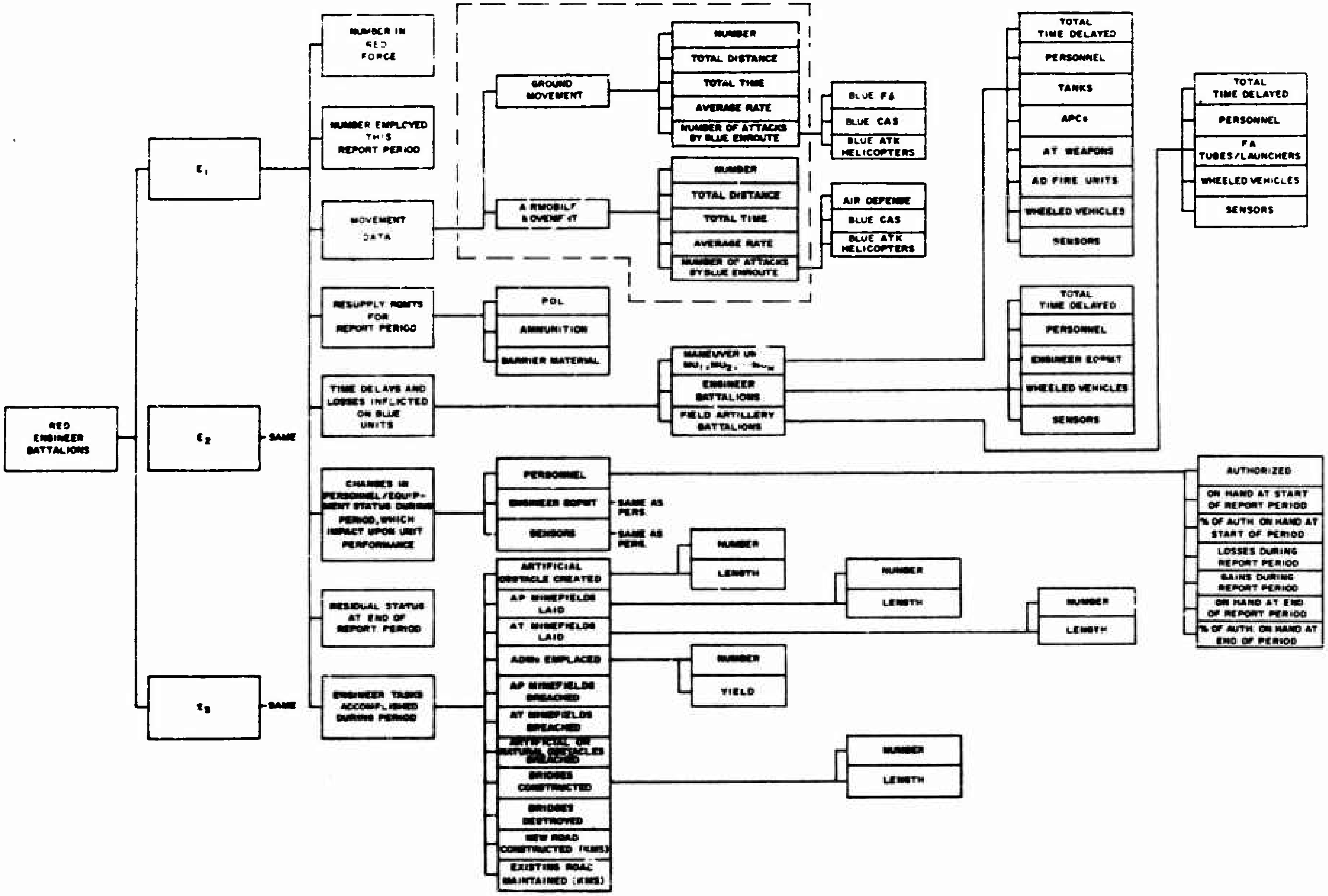


Figure 2-30. Red Engineer Battalions Output Data Array

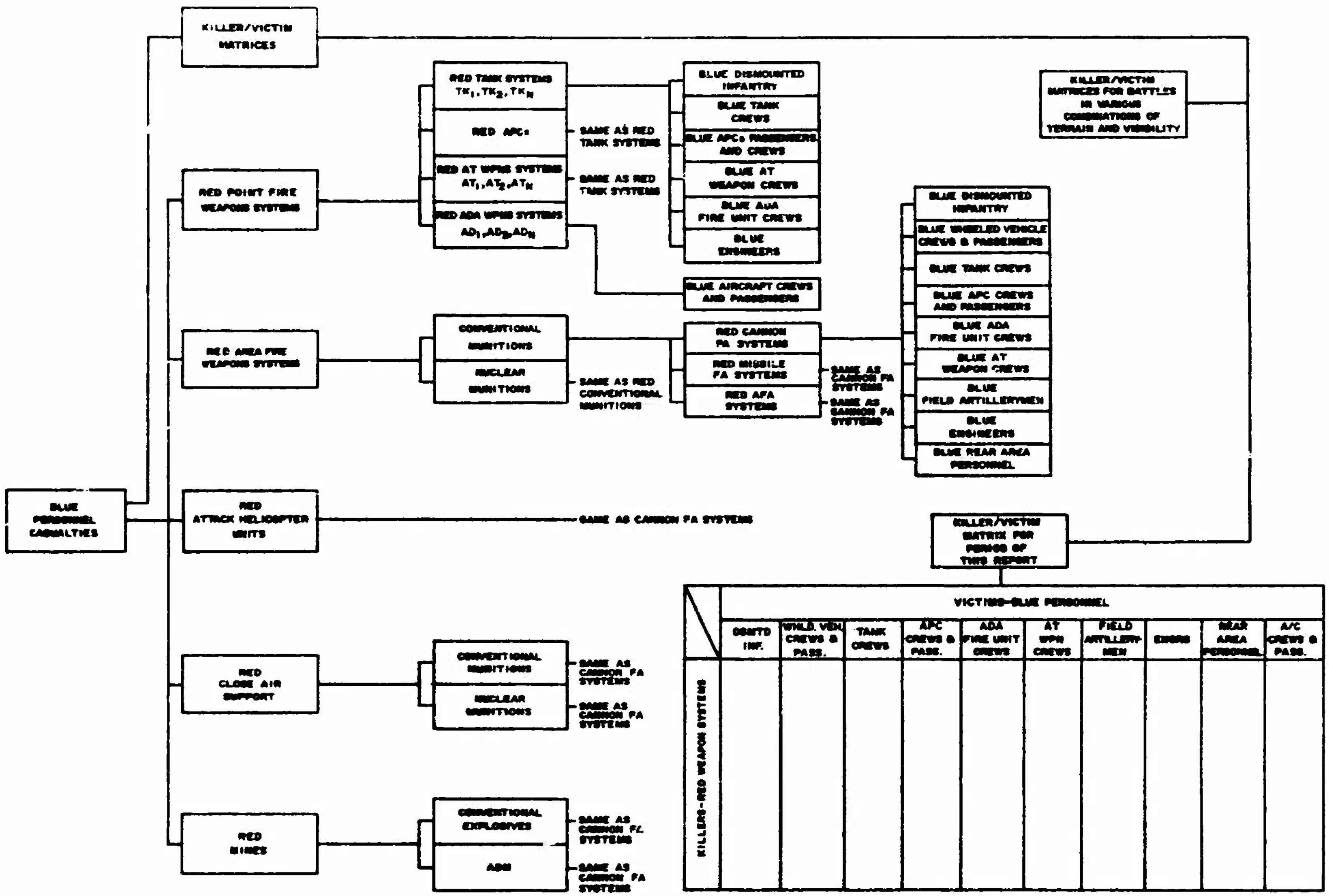


Figure 2-31. Blue Personnel Casualties Output Data Array

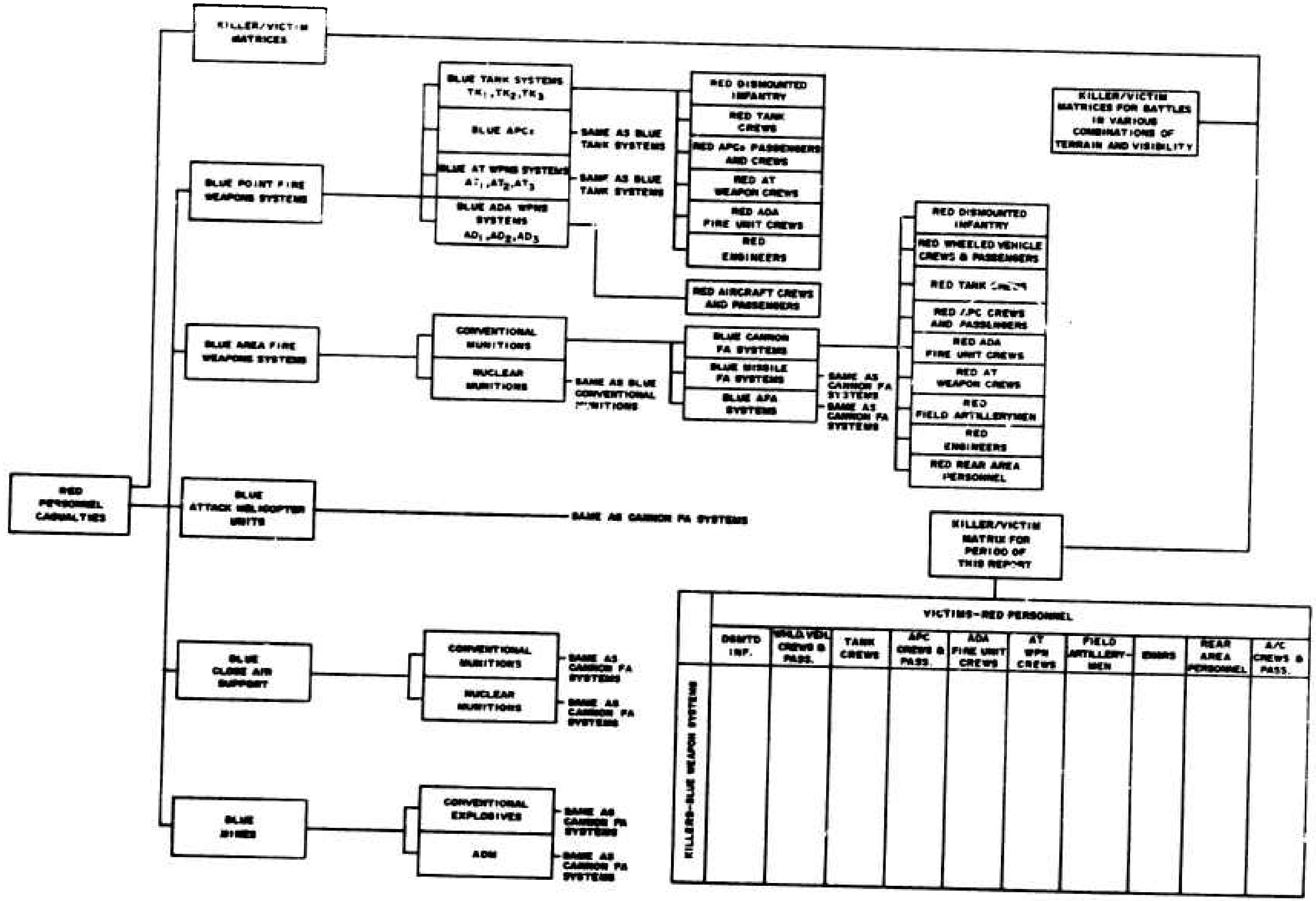


Figure 2-32. Red Personnel Casualties Output Data Array

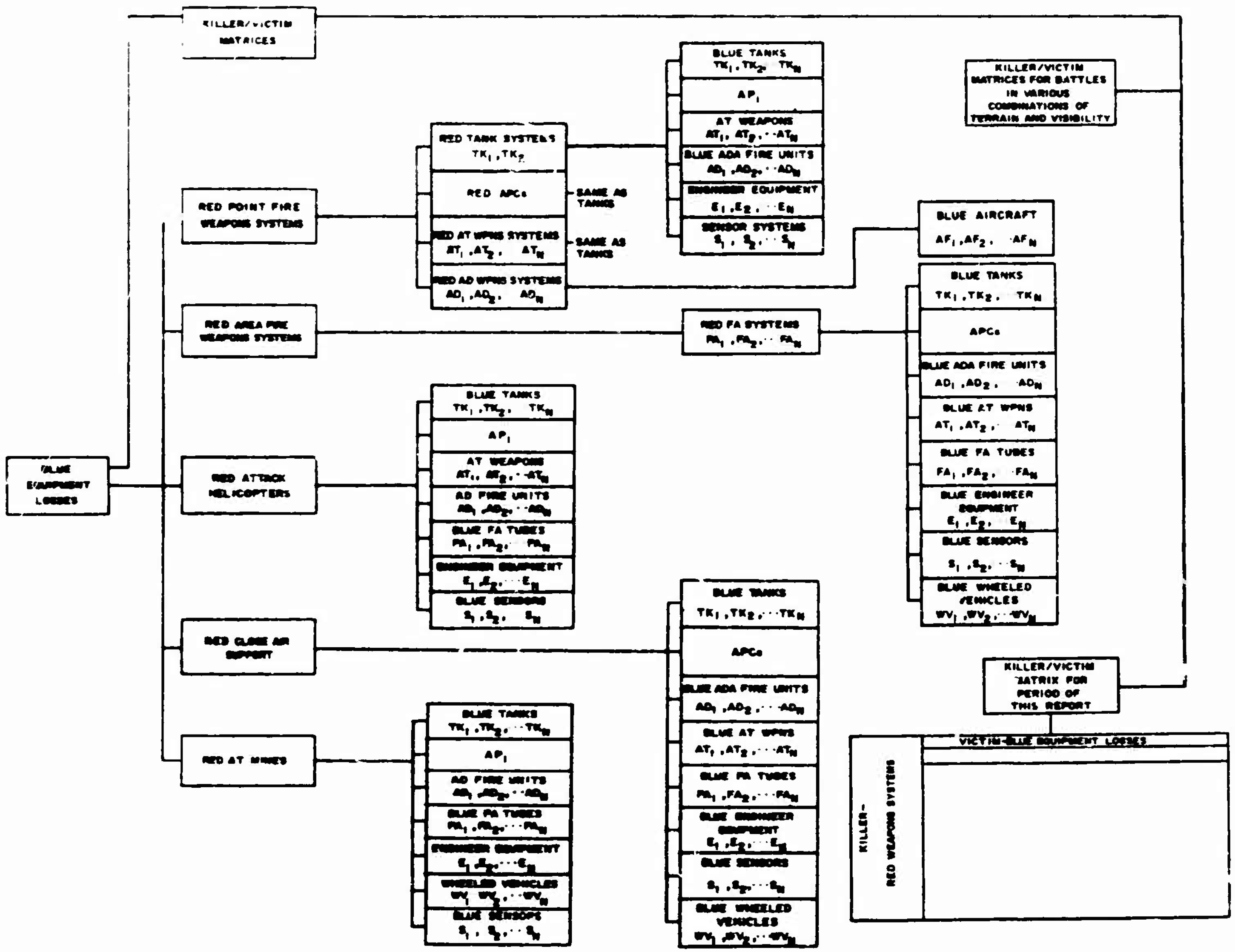


Figure 2-33. Blue Equipment Losses Output Data Array

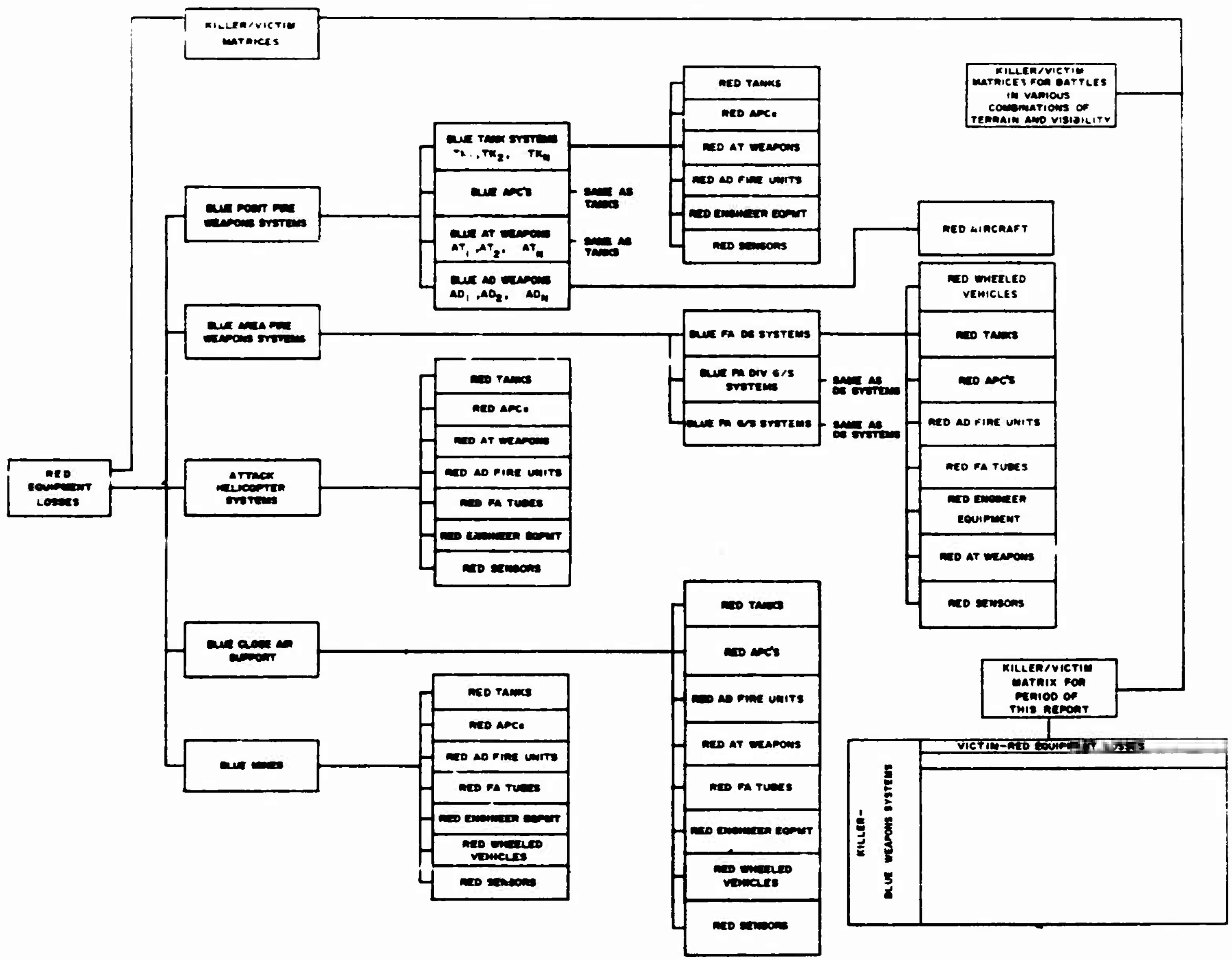


Figure 2-34. Red Equipment Losses Output Data Array

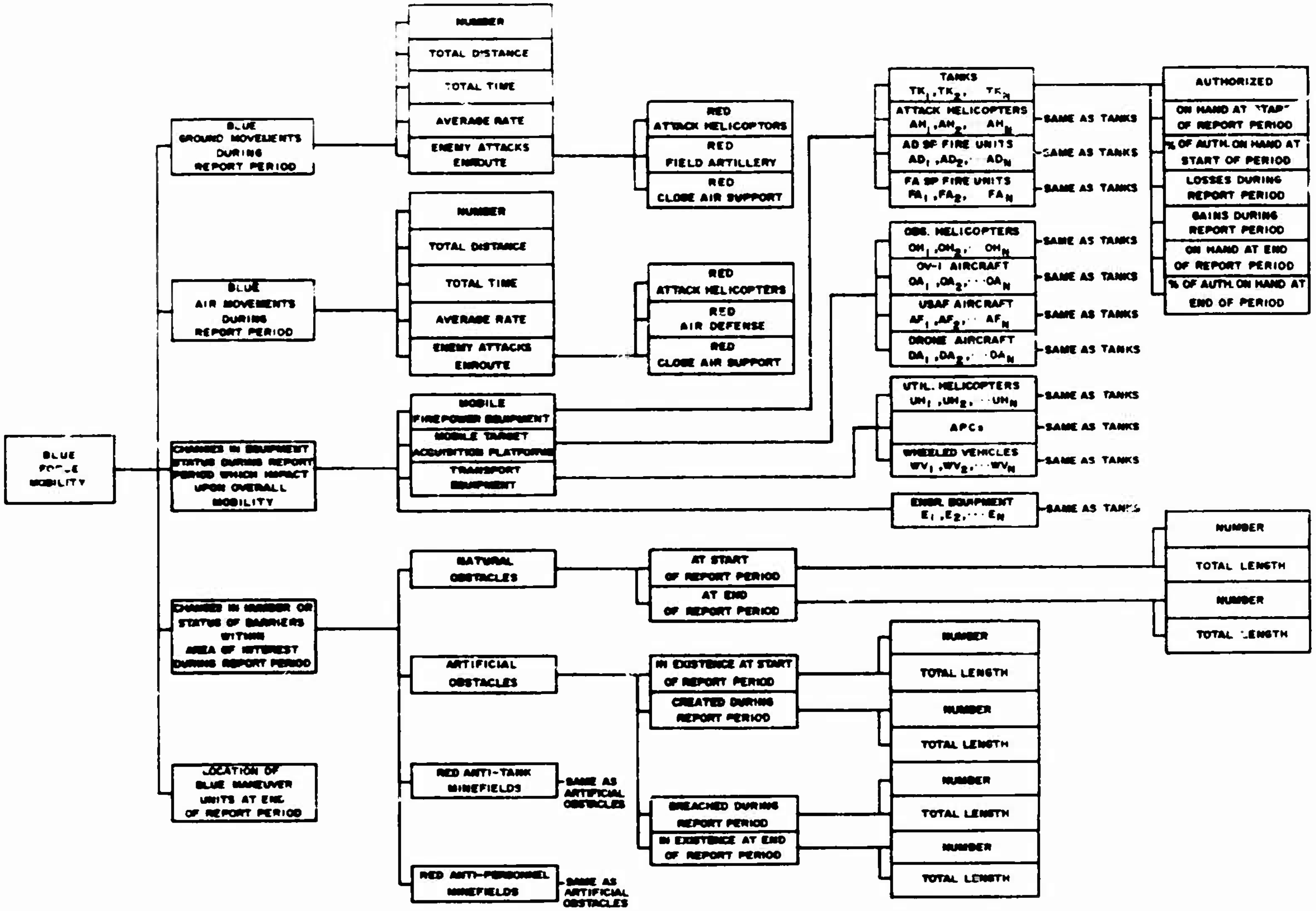


Figure 2-35. Blue Force Mobility Output Data Array

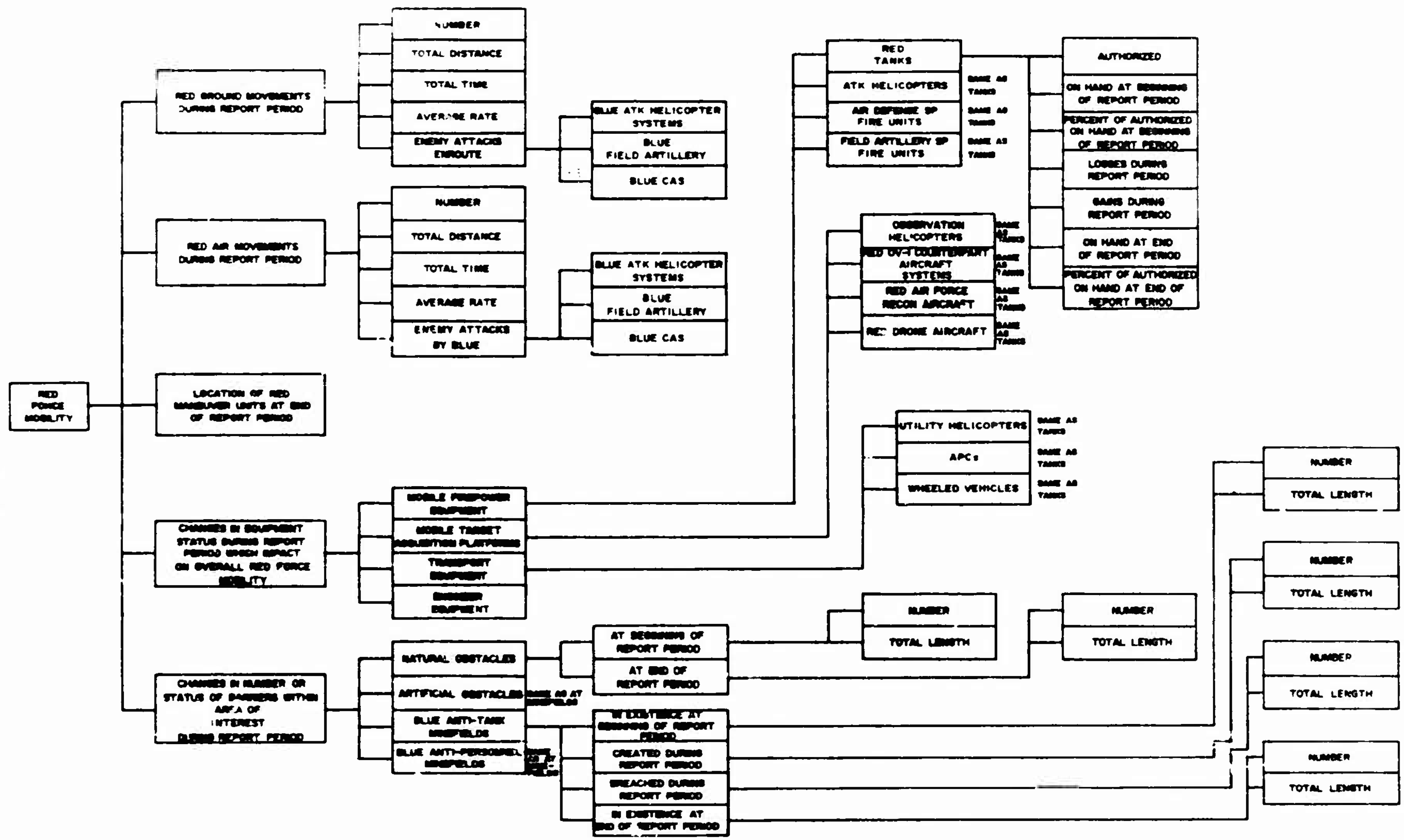
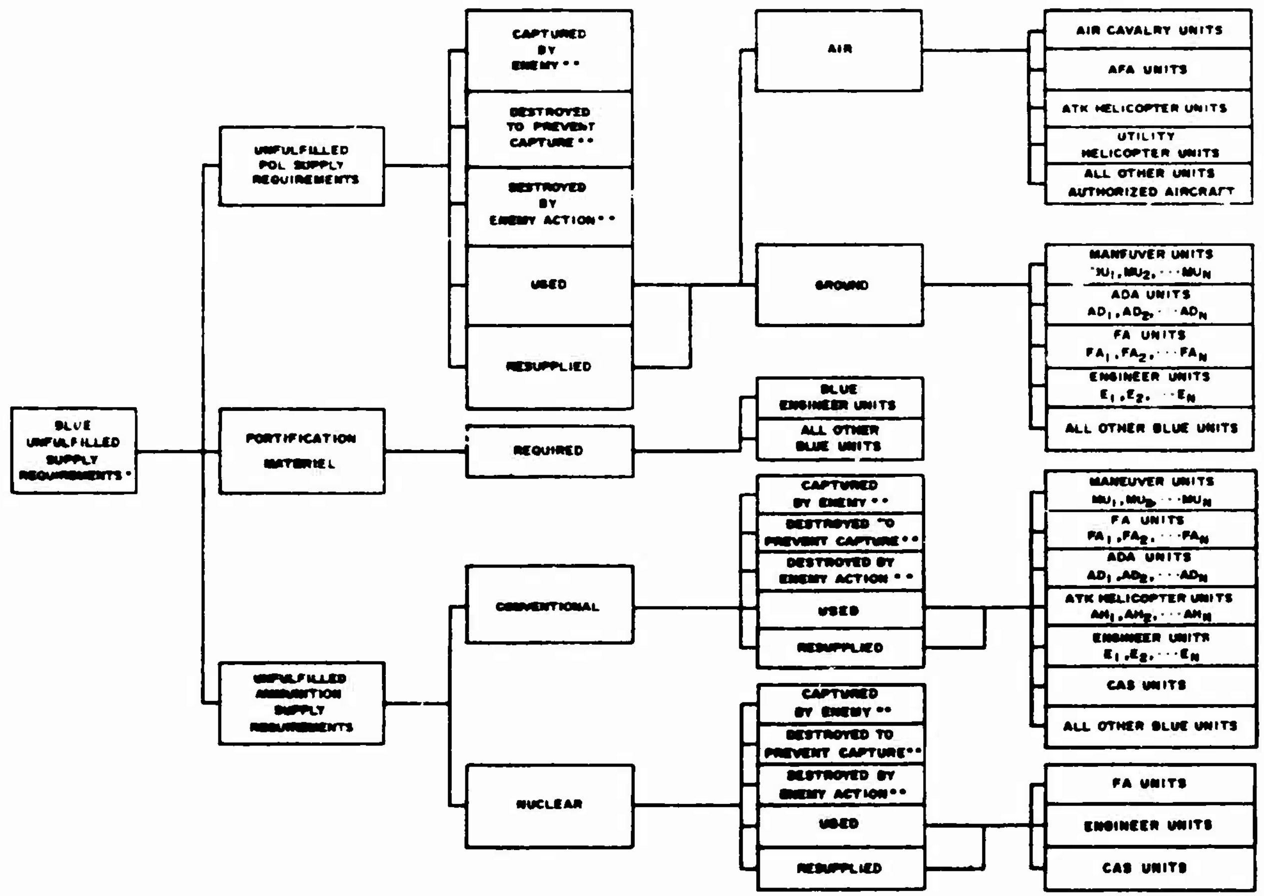


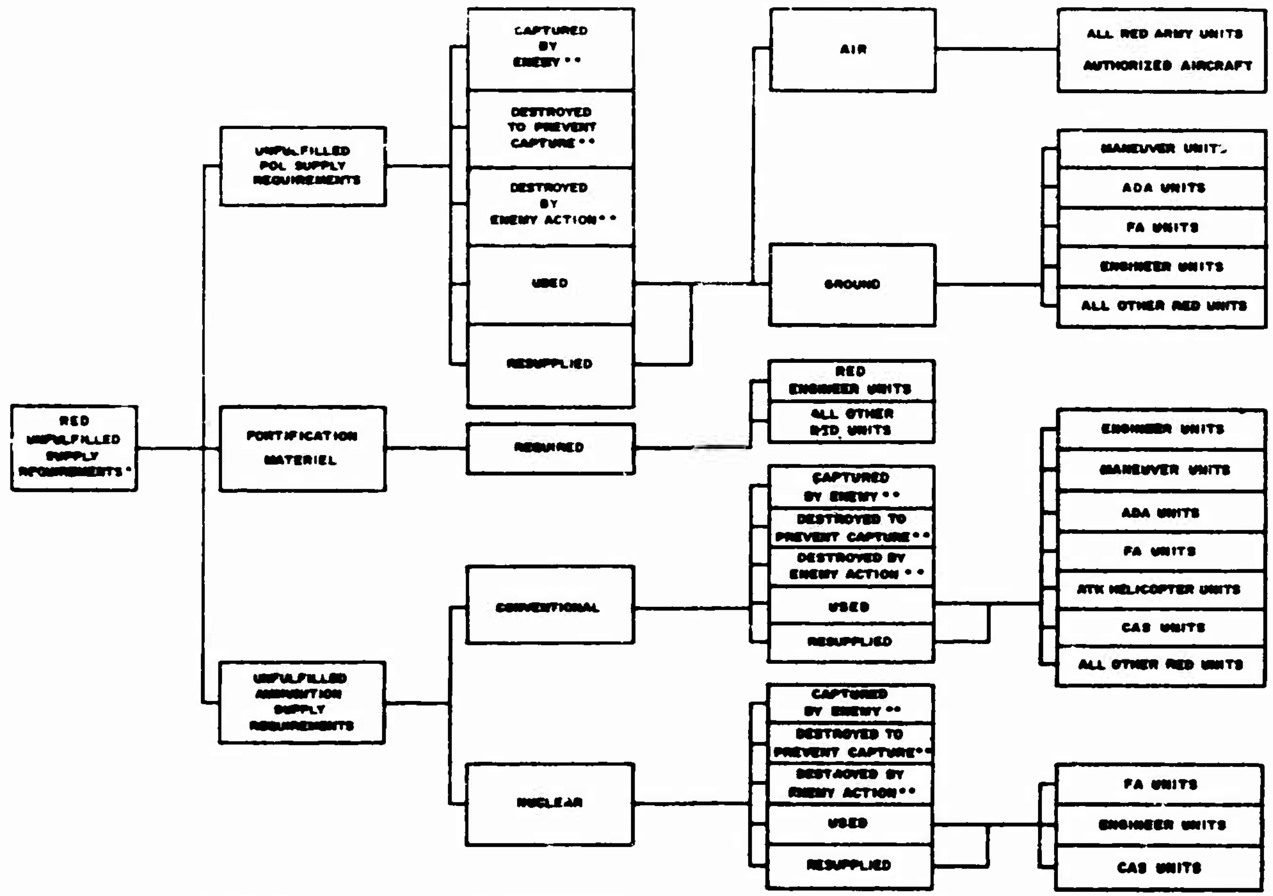
Figure 2-36. Red Force Mobility Output Data Array



*—EXPRESS IN SHORT TONS EXCEPT NUCLEAR ITEMS, WHICH ARE TO BE SHOWN BY NUMBER AND TYPE.

**—EXCLUDES THAT ON BOARD THE FIGHTING VEHICLES.

Figure 2-37. Blue Unfulfilled Supply Requirements



*—EXPRESS IN SHORT TONS EXCEPT NUCLEAR ITEMS, WHICH ARE TO BE SHOWN BY NUMBER AND TYPE.

**—EXCLUDES THAT ON BOARD THE FIGHTING VEHICLES.

Figure 2-38. Red Unfulfilled Supply Requirements

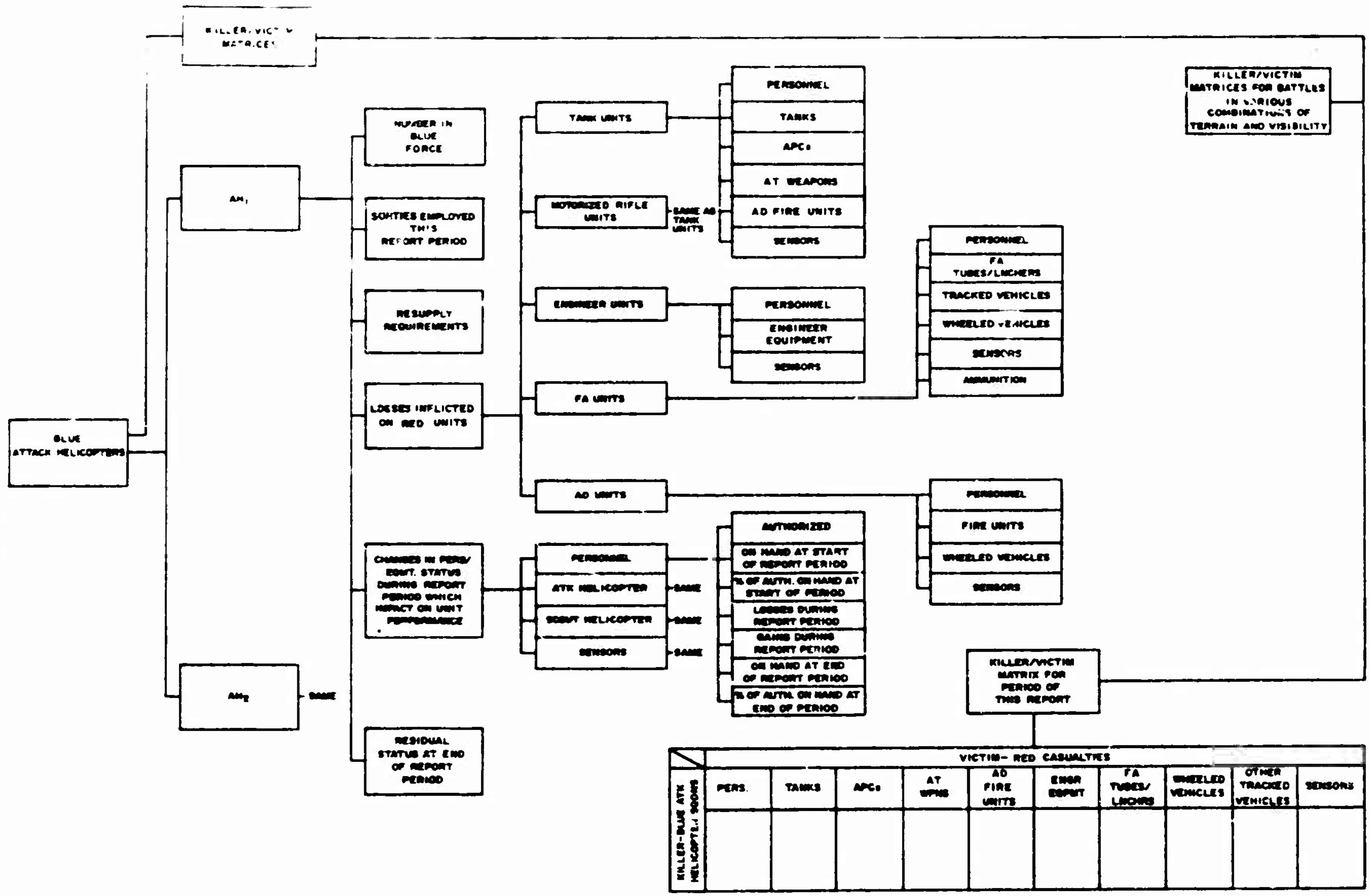


Figure 2-39. Blue Attack Helicopters Output Data Array

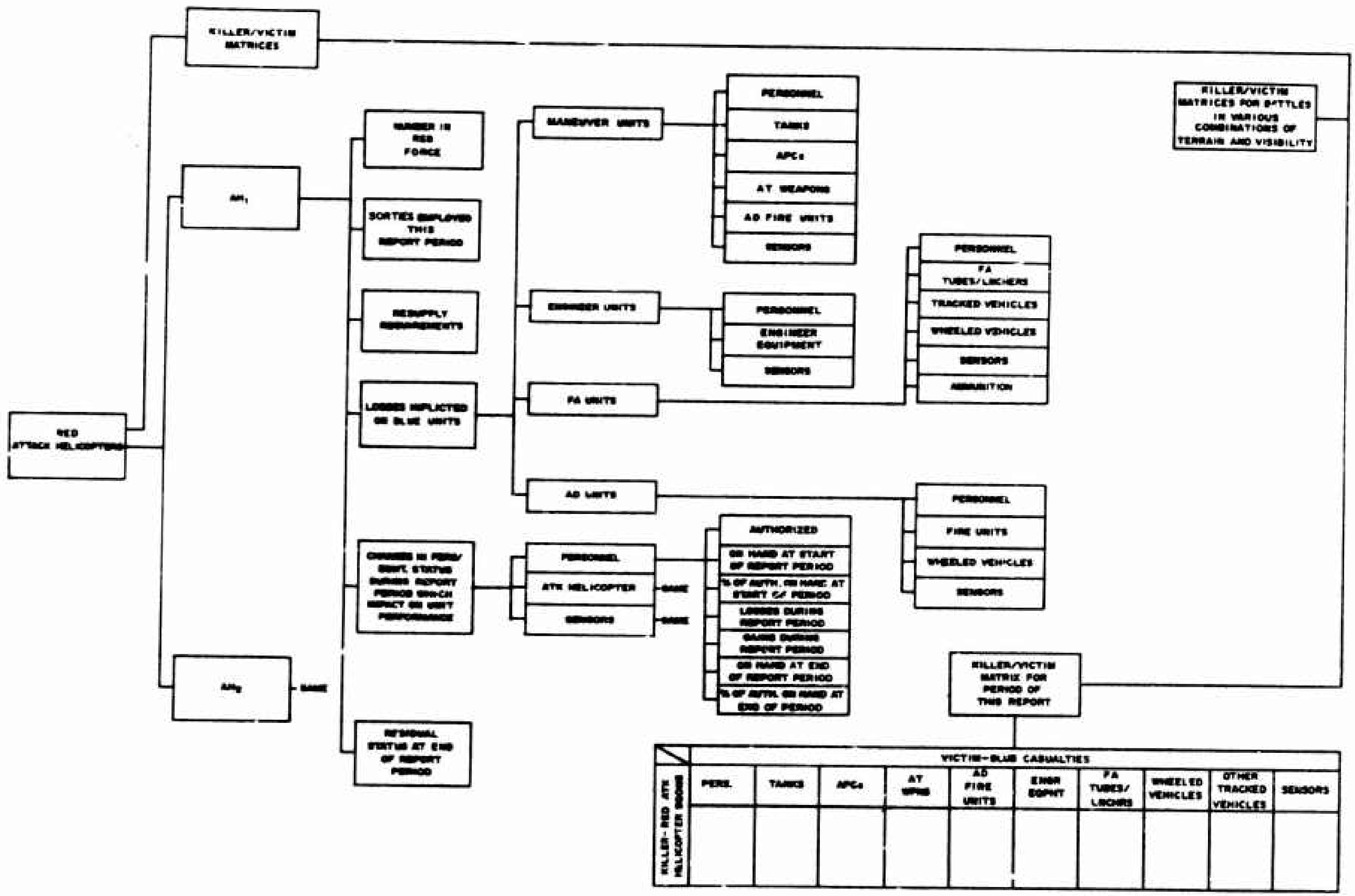


Figure 2-40. Red Attack Helicopters Output Data Array

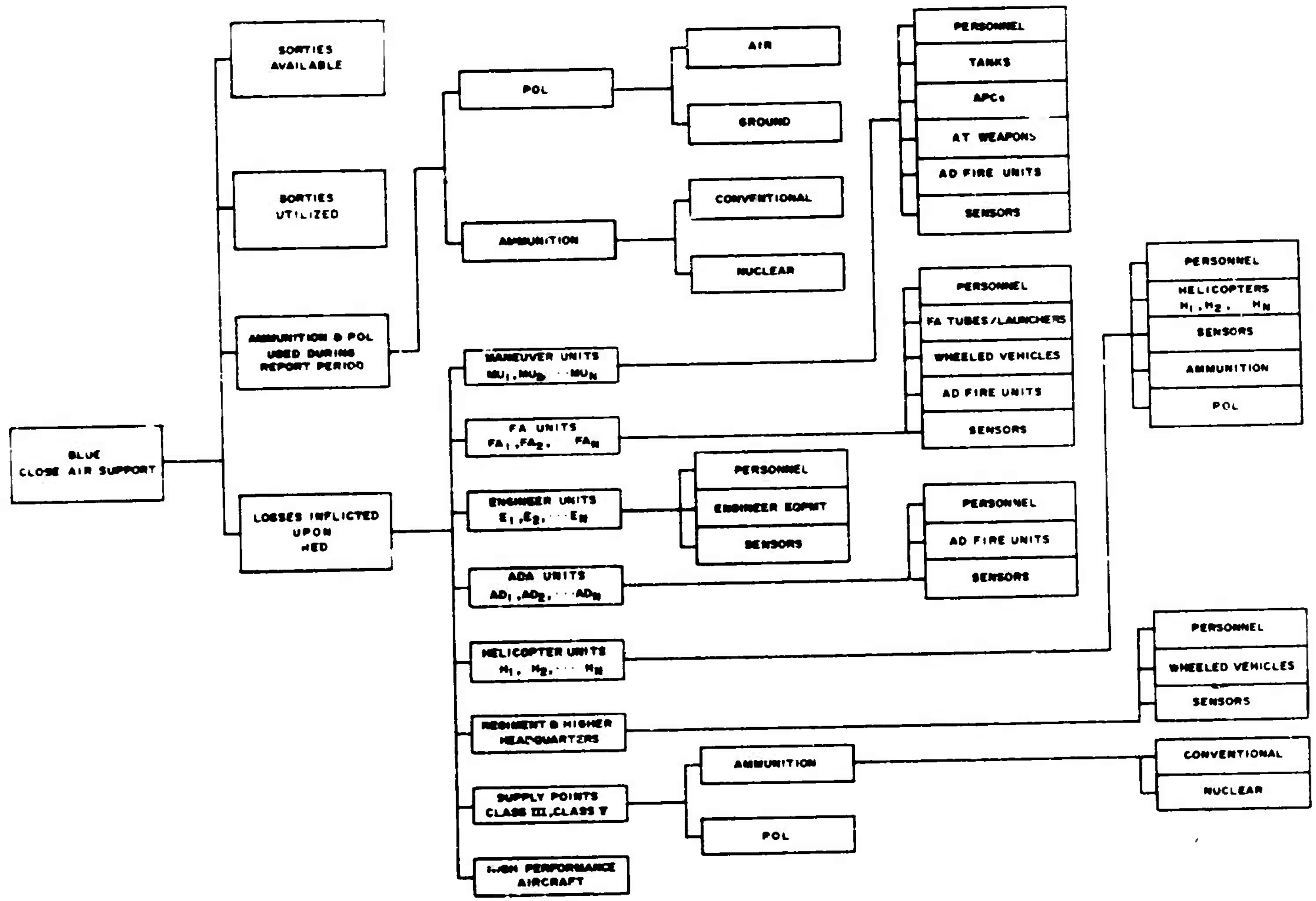


Figure 2-41. Blue Close Air Support Output Data Array

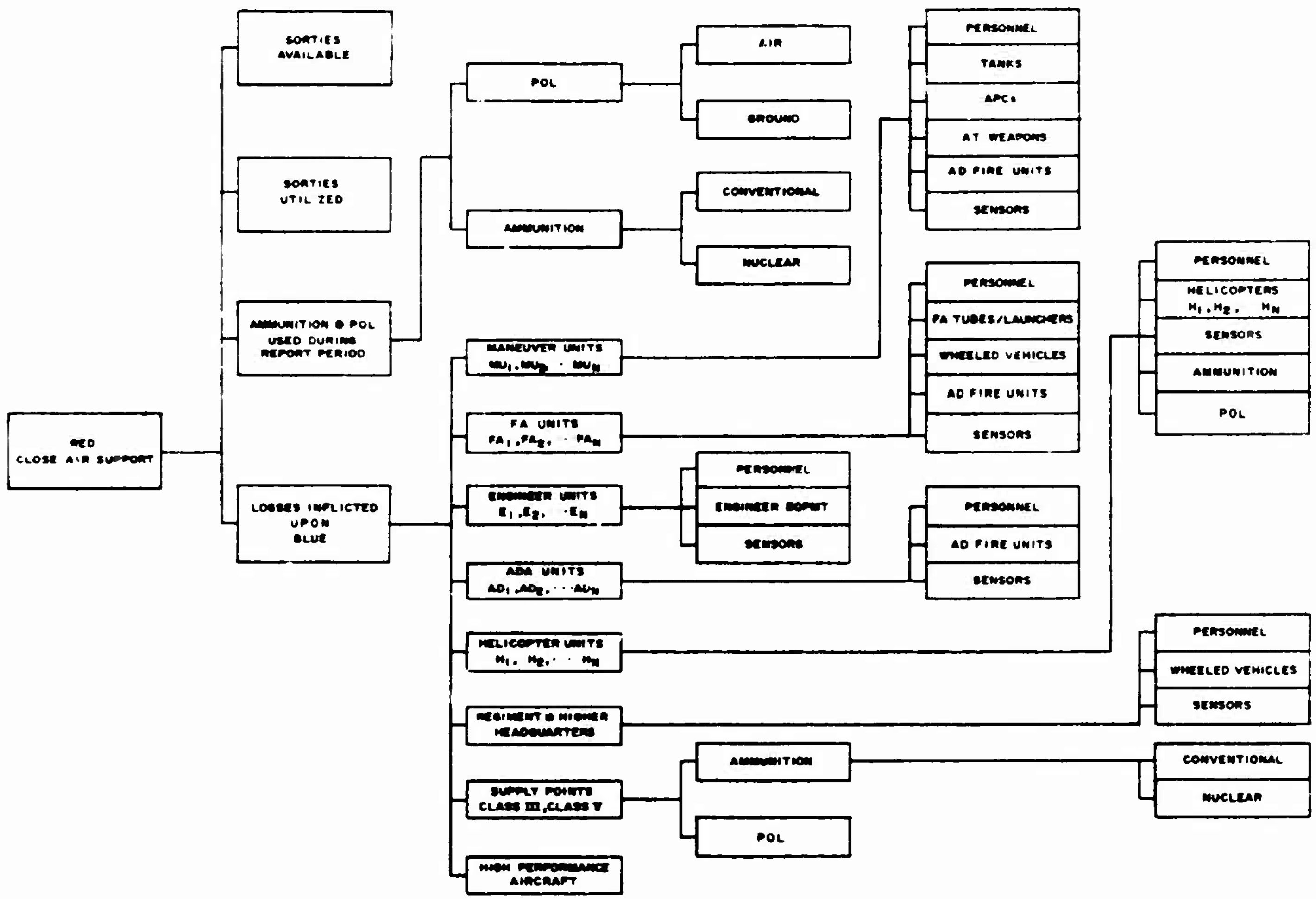


Figure 2-42. Red Close Air Support Outpu⁺ Data Array

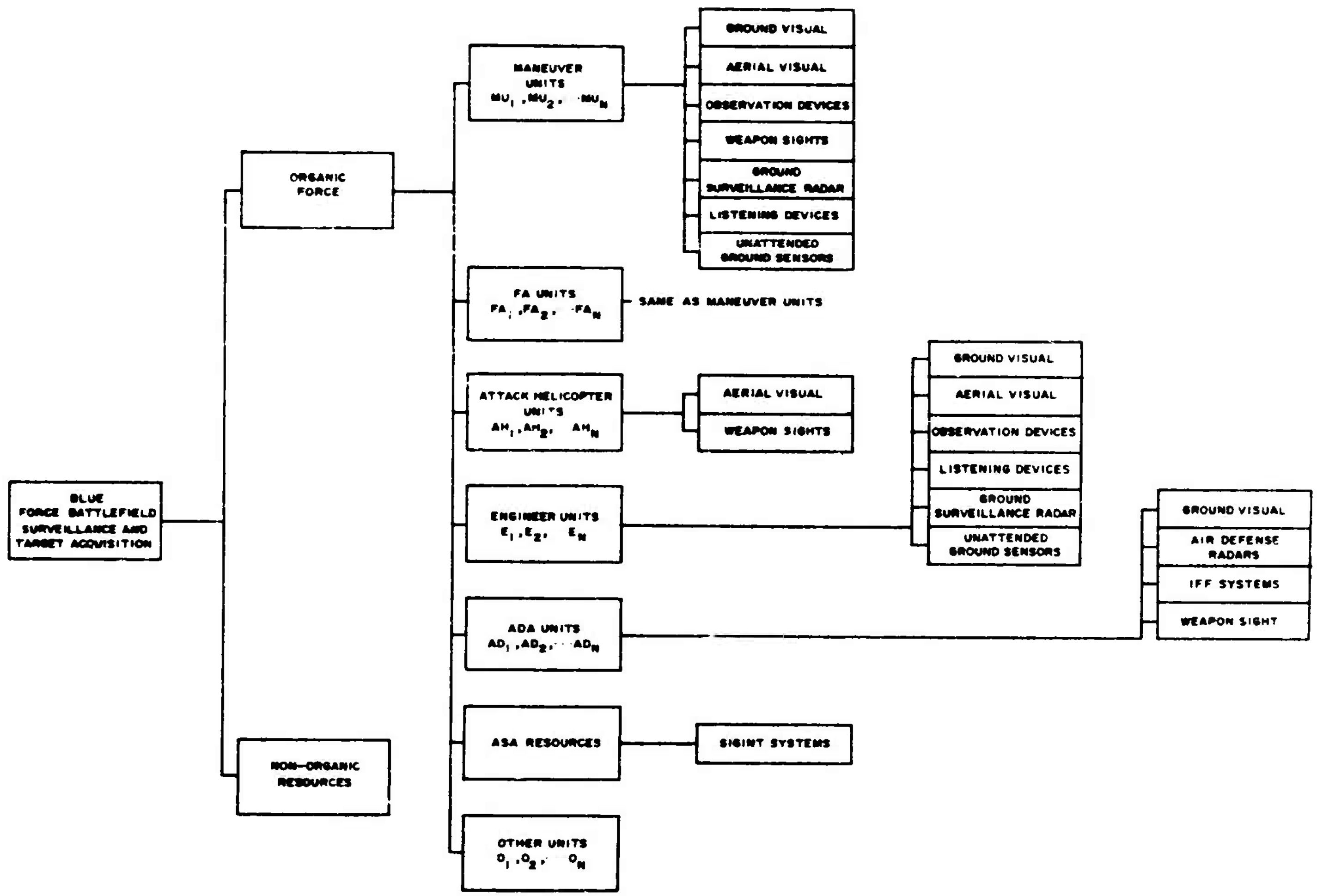


Figure 2-43. Blue Force Battlefield Surveillance and Target Acquisition Output Data Array

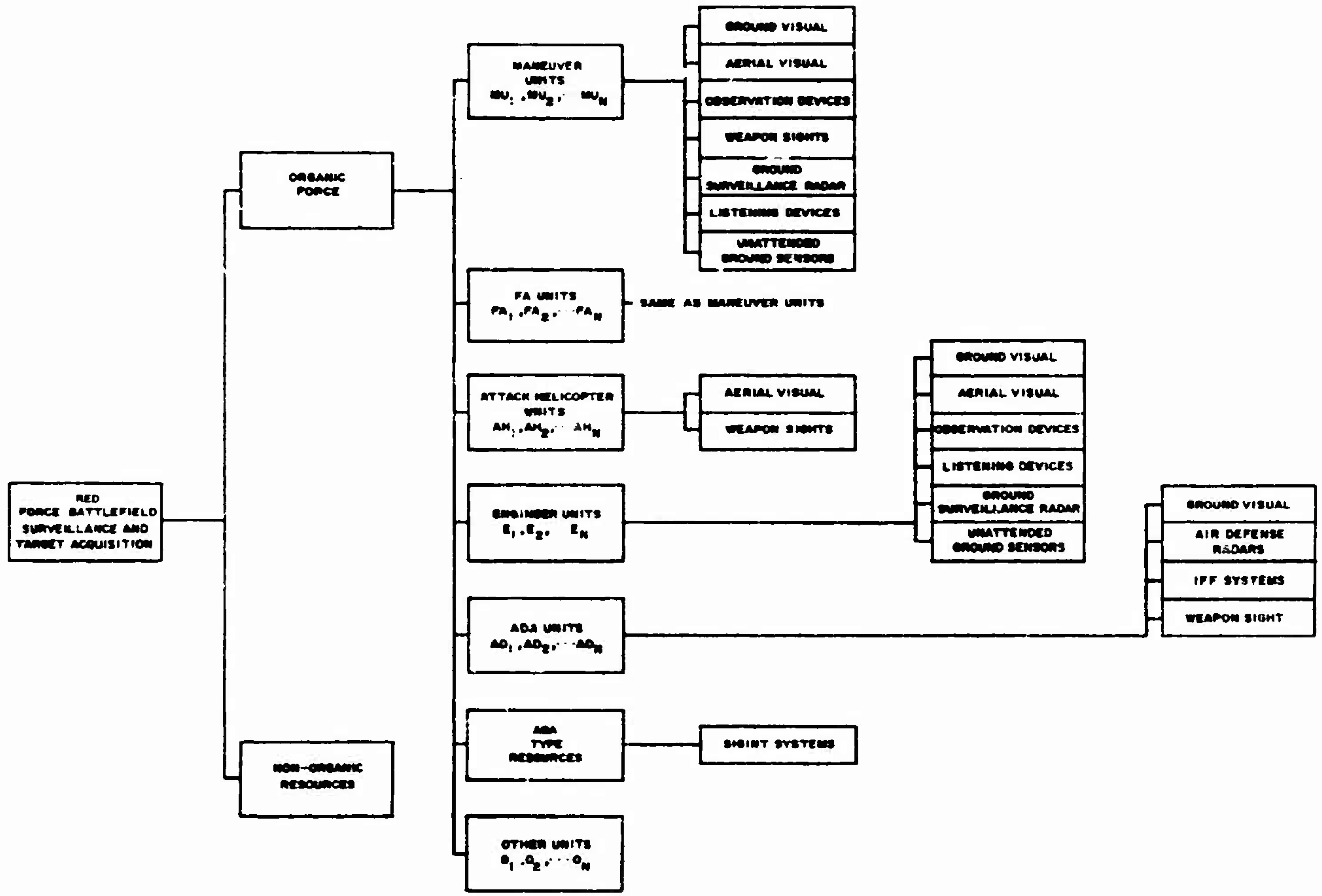


Figure 2-44. Red Force Battlefield Surveillance and Target Acquisition Output Data Array

Blue killers (weapon system)	Red victims																									
	Personnel	Tanks, Tk_1	Tanks, Tk_2	APCs	AD fire units, AD_1	AD fire units, AD_2	FA tubes, FA_1	FA tubes, FA_2	FA tubes, FA_3	Missile launchers	Mortars	AT weapons, AT_1	AT weapons, AT_2	Attack helicopters	Transport helicopters	Obs helicopters	High performance acft	Engineer eqpt, E_1	Engineer eqpt, E_2	Wheeled vehicles	POL*	Conventional ammo*	Nuclear ammo*	Sensors, S_1	Sensors, S_2	Sensors, S_3
Tanks, Tk_1																										
Tanks, Tk_2																										
APCs, AP_1																										
APCs, AP_2																										
AT weapons, AT_1																										
AT weapons, AT_2																										
AT weapons, AT_3																										
AD weapons, AD_1																										
AD weapons, AD_2																										
AD weapons, AD_3																										
AD weapons, AD_4																										
Attack helicopters																										
FA systems, FA_1																										
FA systems, FA_2																										
FA systems, FA_3																										
FA missile systems																										
Mortar systems																										
AFA systems																										
High performance acft																										
Individual weapons																										
Mines, M_1																										
Mines, M_2																										
Total Lost																										
Type Total Lost																										
Total in Force																					NA	NA	NA			
Loss as Percent of Force Total																										

*Destroyed by Blue

Figure 2-45. Blue Firepower Performance Data Array

Section VI. REFERENCES

1. Sidney Siegel, Nonparametric Statistics, McGraw-Hill, New York (1956).

2. Roger E. Kirk, Experimental Design Procedures for the Behavioral Sciences, Brooks/Cole Publishing Company, Belmont, California (1968).

3. Owen L. Davies, The Design and Analysis of Industrial Experiments, Hafner Publishing Company, New York (1956).

4. George W. Snedecor, Statistical Methods, The Iowa State University Press, Ames, Iowa (1962).

5. B. W. Lindgren, Statistical Theory, The MacMillan Company, New York (1968).

6. Thomas S. Ferguson, Mathematical Statistics, A Decision Theoretical Approach, Academic Press, New York (1967).

7. R. Duncan Luce and Howard Raiffa, Games and Decisions, John Wiley and Sons, Inc., New York (1964).

8. Charles H. Kraft and Constance van Eden, A Nonparametric Introduction to Statistics, The MacMillan Company, New York (1968).

9. Rupert G. Miller, Jr., Simultaneous Statistical Inference, McGraw-Hill, New York (1966).

10. P. Nemenyi, Distribution Free Multiple Comparisons, Unpublished doctoral thesis, Princeton University, Princeton, New Jersey (1963).

CHAPTER 3

COMPARISON METHODOLOGY

Section I. INTRODUCTION

1. PURPOSE. This chapter describes a methodology for comparing the combat effectiveness of alternative division forces. The methodology must be used in conjunction with the evaluation methodology described in Chapter 2.

2. ORGANIZATION. This introductory section provides background information relative to the philosophy guiding methodology development. Subsequent sections of this chapter present a step-by-step explanation of the methodology, from the receipt of input data, through subjective and statistical analytical procedures, to a synthesis culminating in a summary of relative force effectiveness. Emphasis is given to a description of the analytical procedures comprising the statistical analysis steps.

3. BACKGROUND:

a. The comparison methodology is intended to be (1) an analytical tool by which any number of division forces may be compared and (2) applicable to data generated by a variety of techniques; e.g., simulations, war gaming, or field tests. It was developed independently of the DIVWAG model but in concert with the evaluation methodology described in Chapter 2. Application of the evaluation methodology produces the data to be used as input in the comparison methodology.

b. The objective of the comparison methodology is to provide a series of standardized techniques that will aid the analyst in relating the combat effectiveness of two or more division forces. A valid comparison of alternative division forces must be conducted by producing data in a format that will permit the examination of the relative effectiveness of the alternative division forces. Additionally, the comparison methodology must permit the comparison of alternative division forces within the bounding parameters of specified mission, threat, and environment. The comparison methodology must also be designed to yield an overall "most effective" force while at the same time accommodating force comparisons requiring more definitive measures of effectiveness such as "most effective against tanks" or "most effective in acquiring targets." To fulfill the objectives the methodology must provide means for:

(1) Identifying the objectives for the comparison of alternative division forces of interest.

(2) Analyzing the composition of the forces to be compared; identifying any differences between threat, doctrine, and/or environmental conditions; and establishing the bounding parameters for the comparison of alternative forces.

(3) Developing an analytical scheme that will produce a quantitative and/or qualitative comparison of alternative division forces based upon the principal measures of effectiveness.

(4) Acquiring data to which the methodology will be applied. These data should relate directly to the measures of effectiveness used to evaluate single division forces.

(5) Presenting the data in a format to facilitate analysis.

(6) Applying the appropriate combination of judgmental (subjective) and mathematical (objective) analyses required to compare forces on both the subordinate unit level and total force level.

Section II. COMPARISON METHODOLOGY DESCRIPTION

4. GENERAL CONCEPT OF COMPARISON METHODOLOGY:

a. The comparison methodology follows a logical sequence of events to arrive at conclusions regarding the relative effectiveness of alternative forces. The process begins with a reexamination of the objectives for the evaluation of each alternative force and other guidance normally provided by the proponent agency, such as any specified essential elements of evaluation.

b. The comparison of alternative forces proceeds through a combination of subjective and statistical analyses similar to those used for the evaluation methodology. Since the evaluation methodology described in the preceding chapter is, in fact, a comparison methodology for constituents of a division, the logical extension of the evaluation methodology is to utilize the same basic procedures to compare division forces. The evaluation of a single force will have produced the overall effectiveness of the single force or forces in various combat activities as well as under various battle treatments; therefore, the next logical step is to apply, as an element of the comparison methodology, the standard statistical tests to the outcomes of each of the single division evaluations.

c. The evaluation analysis output is divided into several data sets for comparative evaluation. For comparison purposes, model output data and the evaluation analyses are available for the alternative division forces. The single force evaluation criteria are important since a valid comparison can be made only after the forces of interest have been gamed so that data exist for their participation in common combat activities. The basis for comparison will be the effectiveness indicators as produced in Chapter 2, Evaluation Methodology. These indicators are sufficiently general so that they can apply to any force structure.

5. COMPARISON METHODOLOGY PROCEDURES. The logic flow for the comparison methodology is shown by Figure 3-1. As indicated therein, the methodology contains four major steps: collection of data, subjective analysis, statistical analysis, and summarization of comparisons.

a. Collection of data includes a reexamination of the objectives for the evaluation of each single force and the establishment of comparison objectives; comparison of the compositions of the forces, threats, doctrine, and environmental conditions; and development of a format for presentation of the output data from the analysis of the alternative division forces.

b. The subjective analysis includes the comparison between the subordinate units/systems of each force using as input data the final set of ranks produced through exercise of the evaluation methodology. This analysis is described in Section III.

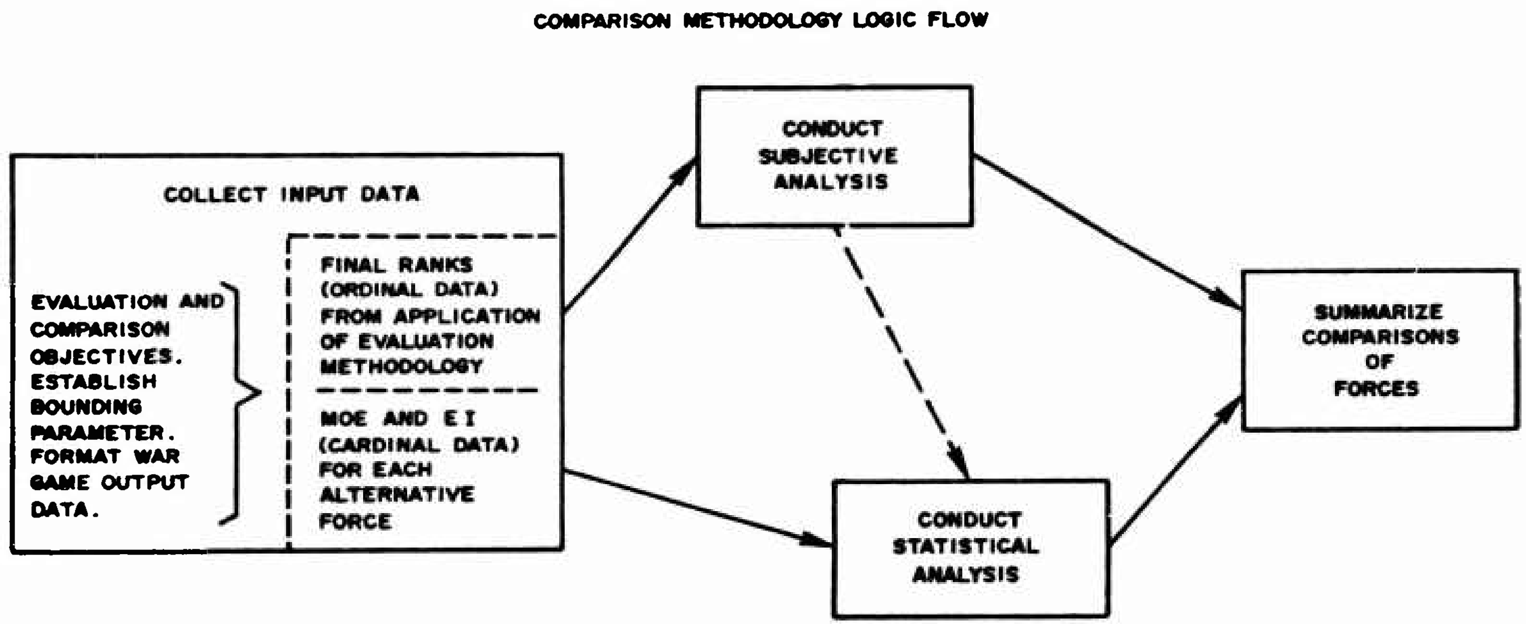


Figure 3-1. Comparison Methodology Logic Flow

c. The statistical analysis includes the comparisons between the forces as a whole and the subordinate units/systems of each force using as input data the model output data translated into measures of effectiveness and effectiveness indicators. This analysis is described in Section IV.

d. The summarization of comparisons includes a synthesis of the analyses conducted and presentation of data for the decision maker. This summarization is described in Section V.

Section III. SUBJECTIVE ANALYSIS

6. INTRODUCTION. This section presents the details of the subjective analysis procedures applied in the comparison methodology. The subjective analysis utilizes the final ranks; i.e., ordinal data, produced through exercise of the evaluation methodology. The ranks were extracted from the data for each force by testing to determine if differences in the game data output between forces were statistically significant. The ranks pertain to type units and type systems; therefore, a direct numerical comparison by inspection of existing subordinate unit ranks indicates the relative value of that unit when considered as a part of the total force. This type comparison is not conclusive with regard to total force effectiveness; however, such comparisons are provided as input to the statistical analysis since they are valuable in allowing the analyst to draw conclusions from which may be formulated null hypotheses for testing through statistical analysis.

7. OUTPUT DATA ARRAYS:

a. A war game, utilizing a model such as DIVWAG, will produce data relating to the strengths and weaknesses of the constituent units and systems of a division force. The application of the evaluation methodology described in Chapter 2 produces both ranks (ordinal data) and subjective analyses for each constituent unit or system of the force and permits comparison between units or systems; i.e., an intraforce comparison. On the other hand, the subjective analysis portion of the comparison methodology, the final output data arrays for type combat units, selected combat support units, and type systems are compared between forces; i.e., an interforce comparison. The data arrays by unit type with systems constant produced in the single force analysis are exemplified in Figure 3-2. A similar data array is produced for each force analyzed using the evaluation methodology.

Unit Type / Activity Type	Tank Heavy Battalion Task Force	Mechanized Heavy Battalion Task Force	Armored Cavalry Squadron	Aerial Artillery Battalion	... N
Attack	1	2	3	3	X
Defense	2	4	4	1	X
Covering Force	3	3	1	4	X
Delay	4	1	2	2	X

Figure 3-2. Example of Final Ranks for Type Unit by Activity Type for One Division Force

b. Similar data arrays are produced in each force analysis by holding the type system constant and ranking the data across combat activity. This type of array is depicted in Figure 3-3.

Activity Type \ System Type	M60A1	XM803	AH-1G	...N
Attack	2	4	1	X
Defense	4	3	4	X
Covering Force	1	1	2	X
Delay	3	2	3	X

Figure 3-3. Example of Final Ranks for Type System by Activity Type for One Division Force

c. These two types of data arrays provide the basic input data for the subjective analysis.

8. CONDUCT OF SUBJECTIVE ANALYSIS. The subjective analysis can be best explained by reference to Figure 3-4, which shows a series of steps providing a logical flow for evaluating the data.

a. Output Arrays by Ranks - Step 1. The exercise of the evaluation methodology for each force produces a series of data arrays by type maneuver unit (e.g., mechanized infantry) and by type systems (e.g., Cobra (AH-1G)) for each force as explained in the preceding paragraph. The first step in the analysis is to group the set of arrays for division forces into common sets, one by type unit and one by type system, preserving the identification of the data set with its associated division force.

b. Select a Unit Type for Each Force - Step 2. The analysis first utilizes the final ranks for type units across combat activities for division forces. One of the type units from the array is selected for analysis.

c. Array Unit Type Ranks for Each Force - Step 3. Once a unit type is selected, the data are arrayed by ranks for each division force and combat activity. An example of such an array is shown in Figure 3-5.

LOGIC FLOW FOR THE SUBJECTIVE ANALYSIS

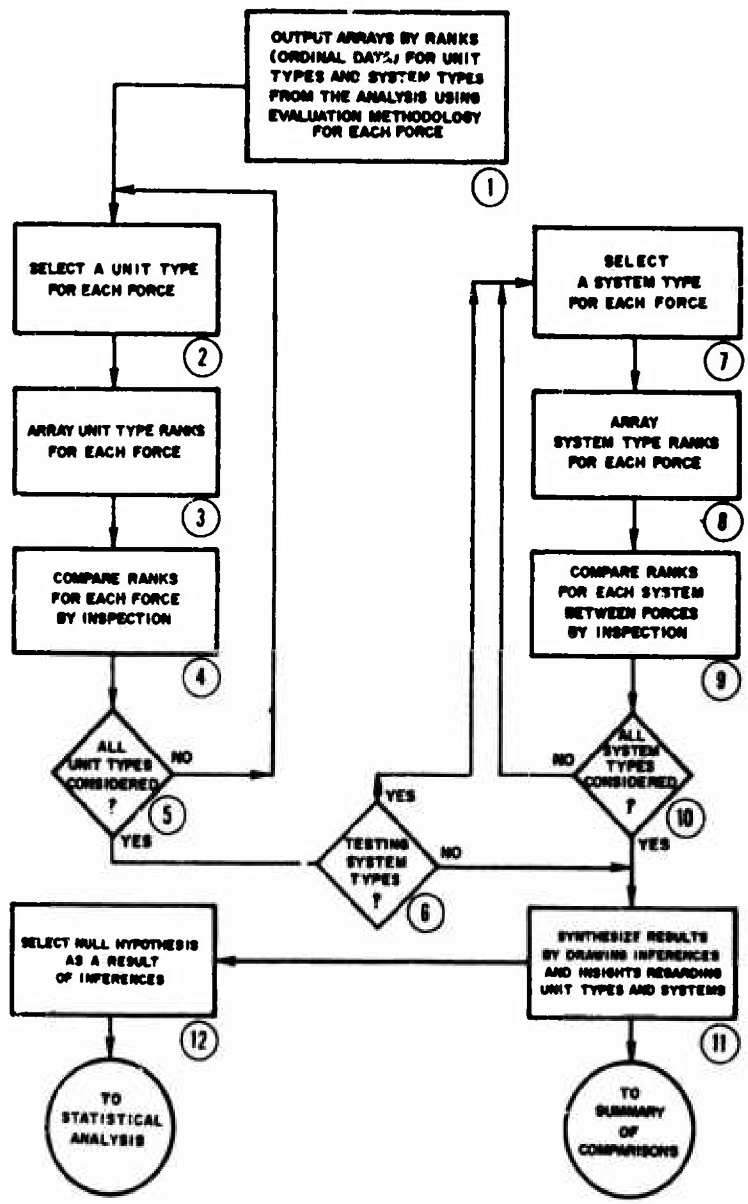


Figure 3-4. Logic Flow for the Subjective Analysis

Activity Type \ Force Type	Tank Heavy Battalion Task Force	
	Division Force 1	Division Force 2
Attack	1	1
Defense	2	4
Covering Force	3	3
Delay	4	2

Figure 3-5. Example Array for One Unit Type Across Combat Activities for Two Division Forces

d. Compare Ranks for Each Force by Inspection - Step 4. If a single unit; e.g., a tank heavy battalion task force, has ranks such as those shown in Figure 3-5, the inference may be made that the tank heavy battalion task forces are equally effective in the attack and equally effective in covering force activities, but that the tank heavy units in Force 1 are more effective in the defense than are those in Force 2. Likewise, the tank heavy unit types in Force 2 are more effective in the delay than are those in Force 1.

e. All Unit Types Considered - Step 5. After completing the analysis for a single type unit, Steps 2 through 4 are repeated until all type units have been inspected. The analysis in Steps 2 through 5 will allow the analyst to gain insights and inferences about the contribution of type units to the combat effectiveness of each force.

f. Testing System Types - Step 6. The purpose of this step is to decide if additional information concerning insight into the relative effectiveness of system types within units is required. If the answer is yes, proceed to Step 7. If the answer is no, then Steps 7 through 10 are bypassed, and results are synthesized (Step 11).

g. Select a System Type for Each Force - Step 7. Insights into the contributions of the systems to each type unit are required; therefore, after exhausting all cases in Steps 2 through 5, a system type is selected within type units across combat activities.

h. Array System Type Ranks for Each Force - Step 8. Once a system type is selected, the data are arrayed by ranks for each division force and combat activity. An example of such an array is shown in Figure 3-6.

Activity Type \ Force Type	M60A1	
	Force 1	Force 2
Attack	2	1
Defend	4	2
Covering Force	3	3
Delay	1	4

Figure 3-6. Example Ordinal Ranks for the M60A1 System Type in Two Different Force Structures Across Combat Activities

i. Compare Ranks for Each System Between Forces by Inspection - Step 9. If a single system type, e.g., M60A1, has ranks such as those depicted in Figure 3-6, the conclusion, by inspection, is that the M60A1 is more effective in the attack when used with Force 2. Likewise, it is more effective in the defense with Force 2 than with Force 1.

j. All System Types Considered ? - Step 10. After completing the analysis for a single system, Steps 6 through 9 are repeated until all system types have been considered.

k. Synthesize Results - Step 11. The results of the analysis are synthesized to determine why the units performed as they did and to evaluate the military significance of any trends and variations detected; thus, unit strengths and weaknesses and, therefore, force structure strengths and weaknesses may be related directly to the differences between forces based on units and systems. This review includes an examination of the battlefield background conditions, the missions assigned to the subordinate units being analyzed, and the orders given to them for the performance of those missions. The synthesis process allows the analyst to make comparisons that have more than a purely judgmental basis. These comparisons are expressed as inferences and insights into the reasons for differences in performance of the two forces. (Insights are defined as intuitive observations not completely supported by available game data.) The review culminates in a summary statement of how the subordinate units and systems performed in the particular combat activity analyzed and why they performed as they did. These summary statements are provided as narrative input to the summary analysis of the comparison of the force.

1. Select Null Hypothesis - Step 12. The results of the subjective analysis are not conclusive with regard to force effectiveness between units and systems. The inferences drawn from the inspection of the ranks will allow the selection of null hypotheses for testing in the statistical analysis. For example, an inference that might be drawn from the analysis is that there is no difference between the effectiveness of the M60A1 tank in Force 1 and Force 2. This is a hypothesis that can be tested readily in the statistical analysis. Each hypothesis postulated is provided as input to the statistical analysis.

Section IV. STATISTICAL ANALYSIS

9. INTRODUCTION. This section presents the details of the statistical analysis procedures applied in the force comparison methodology. The statistical analysis is used to compare (1) subordinate unit performance during a single type combat activity, (2) subordinate unit performance across all combat activities, and (3) aggregated subordinate unit data (brigade and division level) as total forces. This section describes the data output arrays produced from the play of each division force game, the technique whereby data for the alternative forces can be combined, the technique for analysis of subordinate unit performance, and techniques for comparisons of forces at brigade and division levels.

10. OBJECTIVES:

a. The objective of the statistical analysis is to present a coherent and logical mathematical basis for the comparison of the effectiveness of division forces.

b. The statistical analysis within the comparison methodology is an approach to determining comparative and significant strengths and weaknesses between force structures and to assisting an analyst in understanding what occurred during the war game so that proper inferences can be drawn to assist in the decision making process.

11. GAME DATA ARRAYS:

a. The output from the conduct of a war game presents a myriad of data to the analyst. These data must be reduced before logical conclusions and proper inferences can be drawn.

b. The utilization of war gaming data as a technique for making decisions proceeds by properly segregating the data and applying tests to reduce the overwhelming amount of initial data into manageable sets. These sets can be used by the analyst to compare unit effectiveness, system effectiveness, and, eventually, total force effectiveness. This process can be thought of as an effort to make the decision maker's work more tenable.

c. The totality of data from a combat simulation can be visualized using Figure 3-7, where:

i = unit type
j = effectiveness indicator
k = combat activity type
ijk represents the three types.

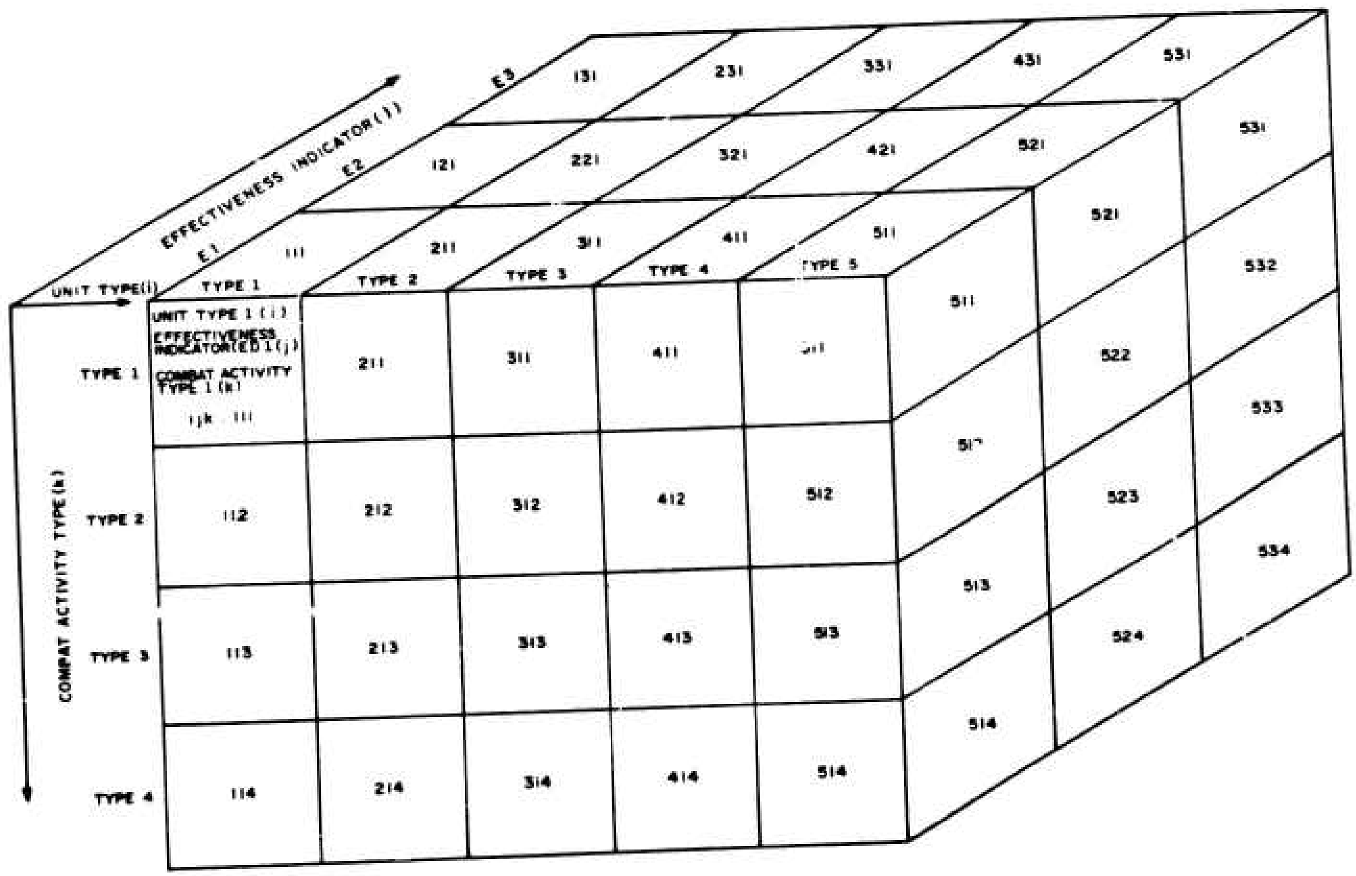


Figure 3-7. Single Force Data Array by Unit Type, Effectiveness Indicator, and Combat Activity

d. Each subscripted block of Figure 3-7 contains engagement data for a selected unit type, a selected activity type, and a selected effectiveness indicator. This form of presentation mentions only unit type; however, every unit type has unit peculiar systemsthat enable that unit to function in its assigned role. The collection of all unit type data infers that system type data will be available for each unit type; thus, for system performance (e.g., firepower systems, intelligence gathering systems, and command service support systems), the words "unit type" may be replaced by "system type." After such a substitution, the methodology may be used to determine the relative impact of different systems on force effectiveness.

e. In a similar vein, one should note that j pertains to effectiveness indicator. Recall (from Chapter 2) that secondary measures of effectiveness pertain to different functions of land combat and that the effectiveness indicators are mathematical entities supporting the secondary MOE. Thus, effectiveness indicators are presented by functional area, and it is possible to consider each functional area independently of any other. Likewise, it is possible to consider an evaluation of the subordinate elements of a single force across all functional areas.

f. As an illustration of these points, an array equivalent to Figure 3-7 is presented in Figure 3-8. Evidently, if such an array is available for each force, a number of direct force comparisons can be made both subjectively and statistically.

12. FILLING THE DATA ARRAY. The process whereby a single game data array is constructed is described herein. Although the array will be filled using computer techniques, and the information of interest for each analysis extracted through a computerized management information system, it should not be inferred that every block of the data array will be filled. This occurs by design because some units (and systems) are designed to function in a single role. Combat service support units, for example, seldom engage the enemy; thus, they have little chance to contribute directly (by creating enemy casualties) to the firepower function of the units active in combat. They may, however, take losses due to area fires and air strikes. When this occurs, their losses will be assessed as would any other unit under fire. Additionally, other units never will engage the enemy due to their deployment. In this case, an array displaying their effectiveness data will have no entry in the firepower function block. Conversely, units responsible for the initial engagement will have few, if any, entries in the combat service support portion of the array. The entries in the array are inserted at different times as a game progresses because the demand for various functions of combat is time dependent.

a. The exact dimensions of an array necessary to accommodate the available evaluation data cannot be predetermined; however, an estimate can be made.

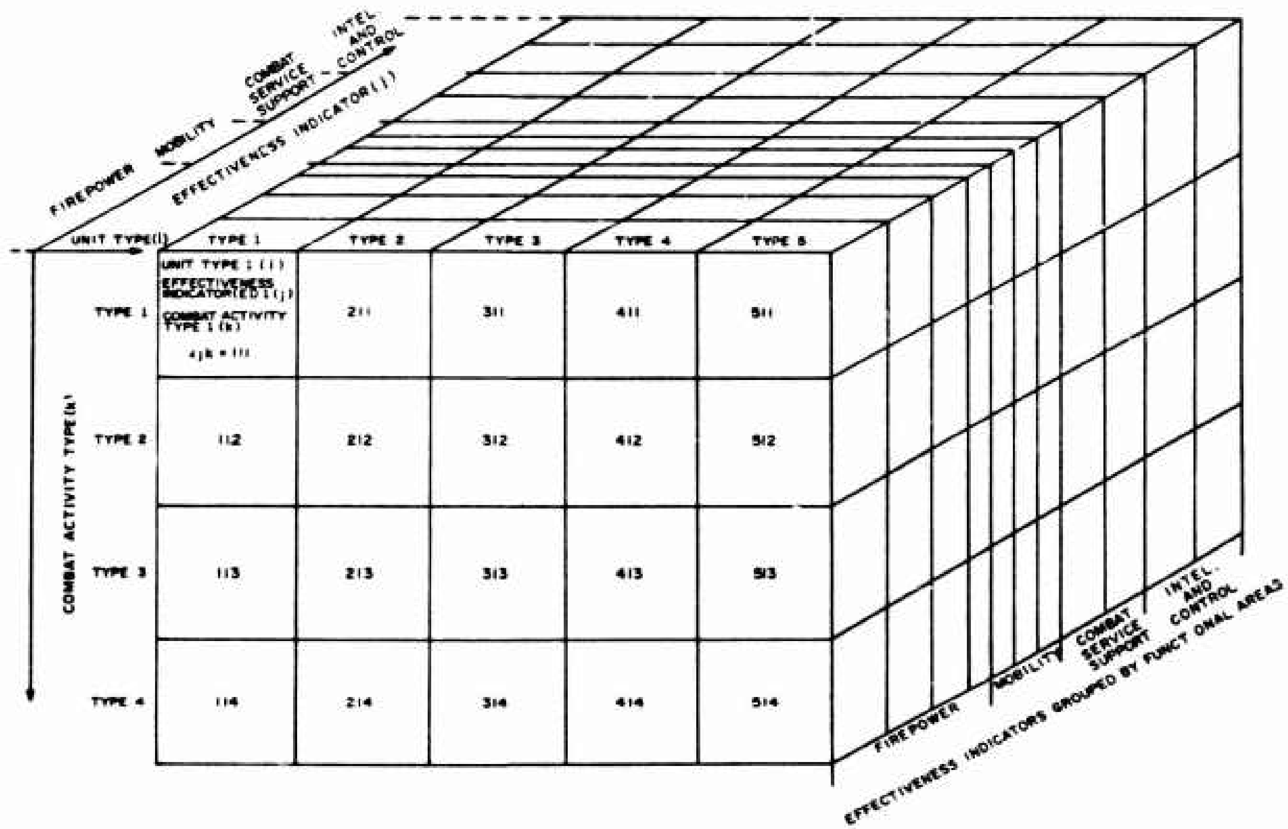


Figure 3-8. Single Force Data Array by Unit Type, Effectiveness Indicators Grouped by Functional Areas, and Combat Activity

(1) A typical 24-hour DIVTAG II output set contained 25 engagements and five distinct weapon mixes. (The 25 engagements were obtained using the rule that each time a reinforcement occurred, a new engagement started.) Using the 15 effectiveness indicators of the firepower function, there are (25) (5) (15) $\approx 2 \times 10^3$ possible data entries per day of combat activity.

(2) The number of data entries may be further expanded across all types of combat activities into, at most, four other groupings, so that ($\approx 2 \times 10^3$) (4) $\approx 8 \times 10^3$ entries could occur per day of simulated activity. For 14 days of activity, this mass of data is not amenable to subjective analysis; however, using the evaluation methodology Phase I (one-way ANOVA), it is possible to meaningfully reduce the data to a manageable size.

b. If the activity is fixed there are 350 data points for a 14-day game period for each effectiveness indicator and weapon mix. The one-way ANOVA reduces these 350 entries to a set of five ranks. Each of five weapon mixes is ranked for each firepower effectiveness indicator considered. In Phase II (two-way ANOVA) the resulting arrav of 75 entries (ranking of five weapon mixes for 15 effectiveness indicators) is reduced to a single set of weapon mix rankings. The final result is approximately a 5×10^3 fold reduction in the data that must be evaluated by the analyst.

c. Although the large size of the original data arrays may appear to be a disadvantage, it is this large size that assures the analyst that the applied statistical procedures have their maximum power and efficiency.

13. COMPARISON TYPES. Three types of comparisons should be conducted: subordinate unit type comparisons by functional area, system type comparisons by functional area, and total force comparisons.

a. Subordinate Unit Type Comparisons by Functional Area. Unit types within each force can be compared between forces to determine their relative effectiveness within each functional area of land combat for each combat activity. This comparison is described in Paragraph 14.

b. Type System Comparisons by Functional Area. Type systems within each force can be compared between forces to determine their relative effectiveness within each functional area of land combat for each combat activity. This comparison is discussed in Paragraph 15.

c. Total Force Comparisons. Direct comparisons between total force structures can be conducted using the procedures and techniques described in Paragraph 16.

14. SUBORDINATE UNIT TYPE COMPARISON BY FUNCTIONAL AREA:

a. Discussion. For this comparison of forces the emphasis is upon the determination of the force structure that is most effective as the functional area of land combat is held constant. Because of differences in force structure, doctrine, threat, and background conditions, there will seldom be the same unit type for each function as the type force is changed. Such a disparity is allowed in the methodology because of the well defined, yet general, effectiveness indicators that have been established for each functional area of combat. The play of the war game will produce a data array for each force as depicted in Figure 3-9. These two arrays are combined for purposes of analysis into a single array where each entry of the new array is coded using an ordered pair (a, i) where a indicates the division force structure and i represents the unit type. This is done so each force will have distinct subordinate units in the analysis.

b. Application of Statistical Analysis. The statistical analyses may be best explained with reference to Figure 3-10, which shows the logic flow of the analysis.

(1) Collect sets of game output data (one set for each force).

(2) Select a functional area.

(3) From each set of game output data extract the subsets of effectiveness indicators by unit type for every combat activity.

(4) Select one activity type.

(5) Select one unit type.

(6) Select one effectiveness indicator to be tested and construct an array of engagement data versus units in the two forces. A typical array is shown in Figure 3-11.

(7) For the array constructed, assign ranks and apply the Kruskal-Wallis one-way ANOVA followed by the Mann Whitney U-Test. This will give a set of ranks such that each unit of this type has a rank for effectiveness indicator 1. This set of ranks is recorded. At this point one null hypothesis (i.e., H_0: All units perform equally well regardless of force and unit structure) has been formulated by subjective inference and tested. Application of the one-way ANOVA will produce a single column of ranks for this effectiveness indicator as shown in Figure 3-12.

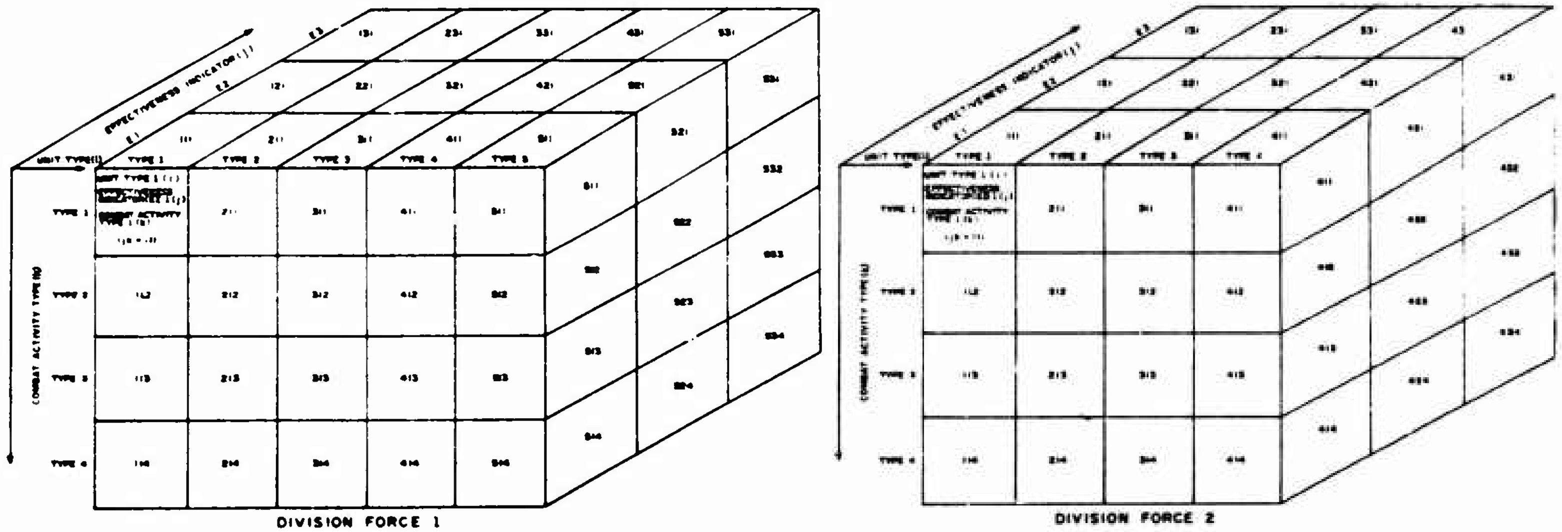


Figure 3-9. Example Data Arrays for Two Division Forces by Unit Type, Effectiveness Indicator, and Combat Activity

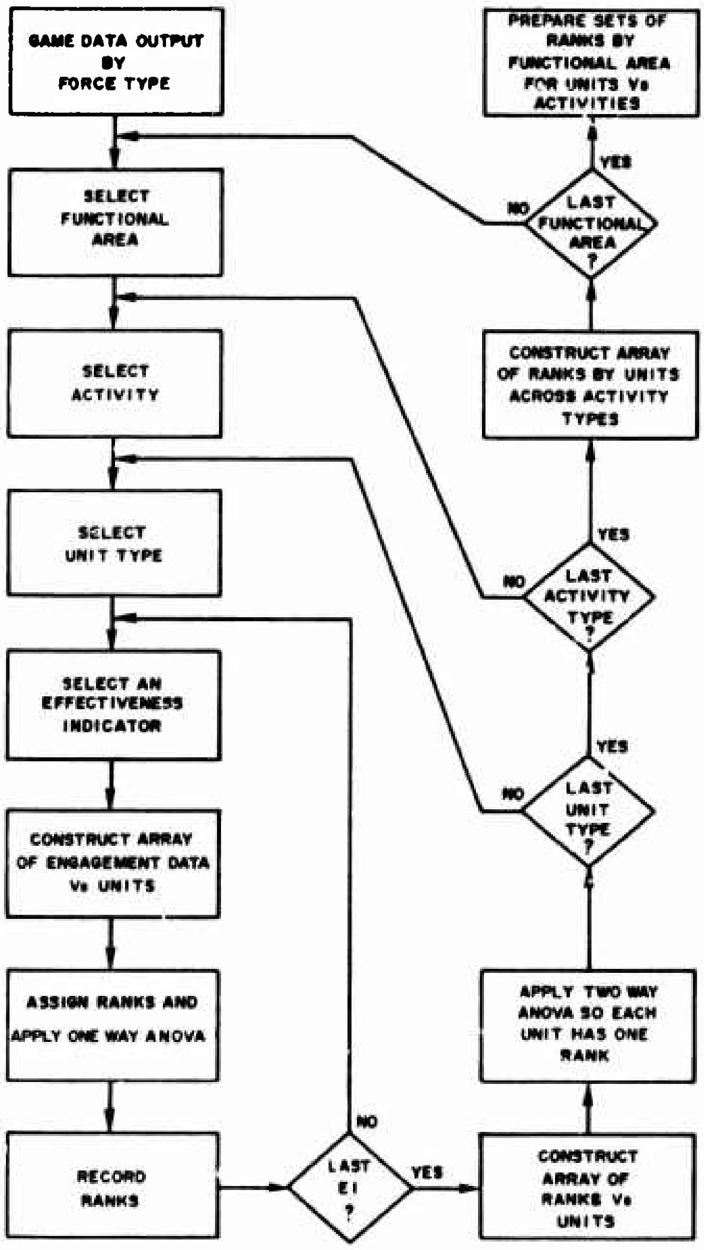


Figure 3-10. Logic Flow for the Statistical Analysis

Engagement Data / Force and Unit (a,i)	ALFA	BRAVO	CHARLIE	DELTA
(1, 1)	Data	Data	Data	Data
(1, 2)	Data	Data	Data	Data
(1, 3)	Data	Data	Data	Data
(1, 4)	Data	Data	Data	Data
(1, 5)	Data	Data	Data	Data
(2, 1)	Data	Data	Data	Data
(2, 2)	Data	Data	Data	Data
(2, 3)	Data	Data	Data	Data
(2, 4)	Data	Data	Data	Data

Figure 3-11. Data Array for Unit Types from Two Forces by Engagement for One Effectiveness Indicator

Force and Unit Type (a, i) \ Effectiveness Indicator	Effectiveness Indicator (1)
(1, 1)	Rank
(1, 2)	Rank
(1, 3)	Rank
(1, 4)	Rank
(1, 5)	Rank
(2, 1)	Rank
(2, 2)	Rank
(2, 3)	Rank
(2, 4)	Rank

Figure 3-12. Ranks for One Effectiveness Indicator for Unit Types

(8) Proceed to effectiveness indicator type 2 for unit type 1 and activity type 1. For the resulting data set (ijk = 121) apply the logic presented in Step 7. Reiterate the procedure for each effectiveness indicator.

(9) After exhausting the set of effectiveness indicators for unit type 1 and activity type 1, collect the ranks into an array of units versus effectiveness indicator. This type array is presented as Figure 3-13.

<table>
<tr><th>Effectiveness Indicator / Force and Unit (a,i)</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th><th>9</th><th>10</th></tr>
<tr><td>(1, 1)</td><td rowspan="9">Ranks for Effectiveness Indicator 1</td><td rowspan="9">Ranks for Effectiveness Indicator 2</td><td rowspan="9">Ranks for Effectiveness Indicator 3</td><td rowspan="9">Ranks for Effectiveness Indicator 4</td><td rowspan="9">Ranks for Effectiveness Indicator 5</td><td rowspan="9">Ranks for Effectiveness Indicator 6</td><td rowspan="9">Ranks for Effectiveness Indicator 7</td><td rowspan="9">Ranks for Effectiveness Indicator 8</td><td rowspan="9">Ranks for Effectiveness Indicator 9</td><td rowspan="9">Ranks for Effectiveness Indicator 10</td></tr>
<tr><td>(1, 2)</td></tr>
<tr><td>(1, 3)</td></tr>
<tr><td>(1, 4)</td></tr>
<tr><td>(1, 5)</td></tr>
<tr><td>(2, 1)</td></tr>
<tr><td>(2, 2)</td></tr>
<tr><td>(2, 3)</td></tr>
<tr><td>(2, 4)</td></tr>
</table>

Figure 3-13. Ranks by Effectiveness Indicator for Unit Types and Fixed Activity.

(10) Apply the Freidman two-way ANOVA and the Mann Whitney two-way test. This will produce a single column of ranks for this activity and function as shown in Figure 3-14.

Force and Unit (a,i) \ Activity	Attack
(1, 1)	7
(1, 2)	6
(1, 3)	9
(1, 4)	1
(1, 5)	2
(2, 1)	3
(2, 2)	5
(2, 3)	4
(2, 4)	8

Figure 3-14. Final Ranks by Unit Type Produced for One Activity

(11) Reiterate through all unit types.

(12) Next, change activity type and reiterate Steps 5 through 11.

(13) Construct an array of unit ranks for each activity type. Figure 3-15 is an example of ranks using one functional area, the activities noted, and the assumption that enough data exist to ensure that each unit participated in each activity. The rank entries are arbitrary and are for purposes of illustration only.

(14) Change functional areas and repeat Steps 3 through 13. The results of this analysis will be an array across units and activities for each functional area as depicted in Figure 3-16. Thus, each secondary measure of effectiveness (i.e., a function of land combat) will have been addressed separately, and a direct comparison of subordinate unit effectiveness can be made between force structures.

Force and Unit (a, i) \ Activity	Attack	Defend	Withdraw/ Delay	Covering Force
(1, 1)	7	2	1	1
(1, 2)	6	6	2	3
(1, 3)	9	3	4	2
(1, 4)	1	4	3	4
(1, 5)	2	5	6	5
(2, 1)	3	1	7	6
(2, 2)	5	7	5	7
(2, 3)	4	9	8	8
(2, 4)	8	6	9	9

Figure 3-15. Ranks for One Function Across All Participating Units and All Activities

Intelligence and Control

Combat Service Support

Mobility

Firepower

Force and Unit (a,i) \ Activity Type	Attack	Defend	Withdraw/ Delay	Covering Force
(1, 1)	7	2	1	1
(1, 2)	6	6	2	3
(1, 3)	9	3	4	2
(1, 4)	1	4	3	4
(1, 5)	2	5	6	5
(2, 1)	3	1	7	6
(2, 2)	5	7	5	7
(2, 3)	4	9	8	8
(2, 4)	8	8	9	9

Figure 3-16. Ranks for Unit Types by Combat Activity for Functional Area

15. TYPE SYSTEM COMPARISONS BY FUNCTIONAL AREA:

a. The play of the war game will produce data arrayed by system types, effectiveness indicators, and combat activity. Examples of the arrays for two division forces are presented in Figure 3-17. The difference in this figure and the example given in Figure 3-9 is that the i axis now contains data pertaining to types of systems. The two arrays in Figure 3-17 are combined for purposes of analysis into a single array where each entry of the new array is coded using an ordered pair (a, i) where a indicates the division force structure and i represents the system type. The comparison of two forces is conducted using the logic flow in Figure 3-10 and the procedures in Paragraph 14b by changing the words "unit type" to "system type." The result of this analysis will be an array of system types versus combat activities for each of the functional areas of land combat. These arrays are presented in Figure 3-18.

Intelligence and Control

Combat Service Support

Mobility

Firepower

Force and System Type (a,i) \ Activity	Attack	Defend	Withdraw/ Delay	Covering Force
(1, 1)	7	2	1	1
(1, 2)	6	6	2	3
(1, 3)	9	3	4	2
(1, 4)	1	4	3	4
(1, 5)	2	5	6	5
(2, 1)	3	1	7	6
(2, 2)	5	7	5	7
(2, 3)	4	9	8	8
(2, 4)	8	8	9	9

Figure 3-18. Ranks of System Types by Combat Activity for Each Functional Area

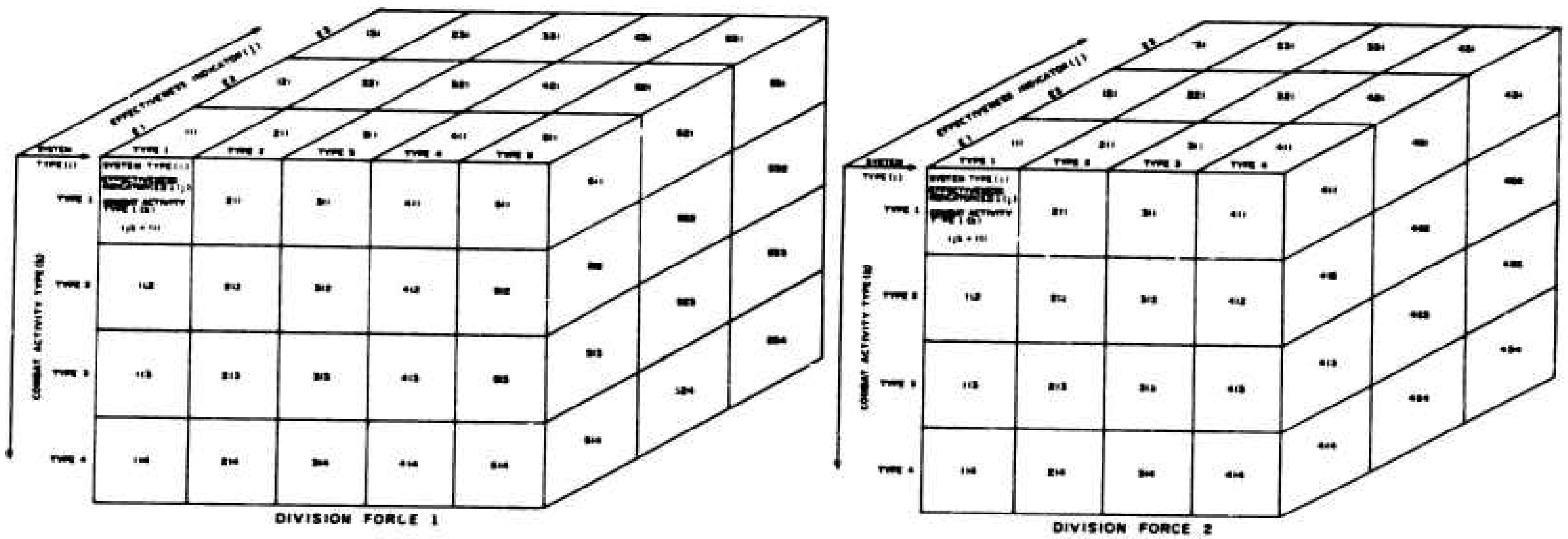


Figure 3-17. Example Data Arrays for Two Division Forces by System Type, Effectiveness Indicator, and Combat Activity

b. This analysis will provide a test of the null hypothesis concerning systems postulated in the subjective analyses and a summary which includes:

(1) A listing of the identified strengths and weaknesses with an assessment of their significance:

(a) Where the strengths and weaknesses are peculiar to a single functional area.

(b) Where the strengths and weaknesses extend across two or more functional areas.

(2) A comparison of all insights derived from the comparison of systems by combat activity, with consistencies and inconsistencies identified.

16. SINGLE BRIGADE FORCE COMPARISONS:

a. The statistical analysis presented to this point has considered examples wherein the battalion was the resolution element, or smallest subordinate unit type worthy of consideration. The methodology is not restricted to this level of aggregation and may be applied to any higher level of resolution such as companies and platoons. Similarly, the methodology may be utilized at lower resolution levels such as brigade and division force levels. This paragraph treats brigade level comparisons both within a single force and across forces. The treatment across forces considers only two forces; however, it may be generalized to consider any number of force structures.

b. Consider a single battle common to both forces, which occurs during a fixed amount of time, and assume, for example purposes, that the forces engage a common threat. The scale in time will be carried in the computer as discrete, yet ordered, time increments. Conceptually, the data in this file may be represented by a single line; and an X on that line represents a change in events such as reinforcement, withdrawal, loss of all of a major weapons system, or any other significant event. The interval $t=0$ to $t=t_1$ is the first increment, $t=t_1$ to $t=t_2$ is the second increment, etc., as shown in Figure 3-19. Incremental time phasing of this type is possible for each battalion level encounter. Using a situation similar to that analyzed in Chapter 2, there are four battalion level encounters constituting two brigade level encounters. Thus, four files of significant times must be kept. The data set for the two brigades and the division may be uniquely related to these files.

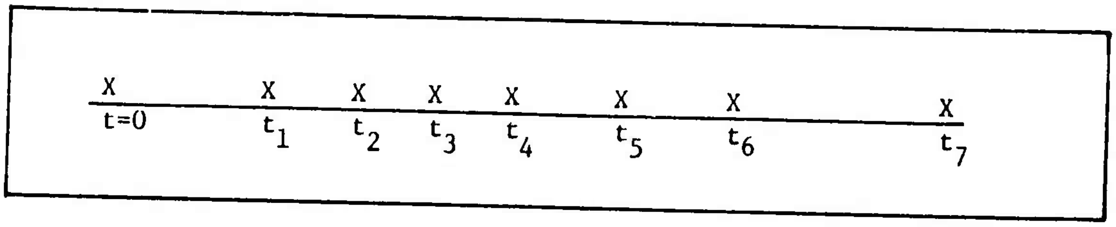


Figure 3-19. Conceptual Time Increment Line

c. Data on file for battalions may be used to construct a data file for each brigade. The technique is one whereby the two files are combined as will be illustrated for Brigade 1 of Force 1. Consider Blue personnel casualties and their time dependence within each battalion. Since two files are of interest there will be two data sets, as shown in Figure 3-20. In this figure ABC (t_i, t_j) is a notation used for convenience. The A implies that battle ALFA is important, BC indicates that Blue casualties are under consideration, and the pair (t_i, t_j) indicates the time interval for the battle of interest. Similarly, BBC_4 indicates that in the fourth time increment of battle BRAVO the value of Blue casualties is of interest.

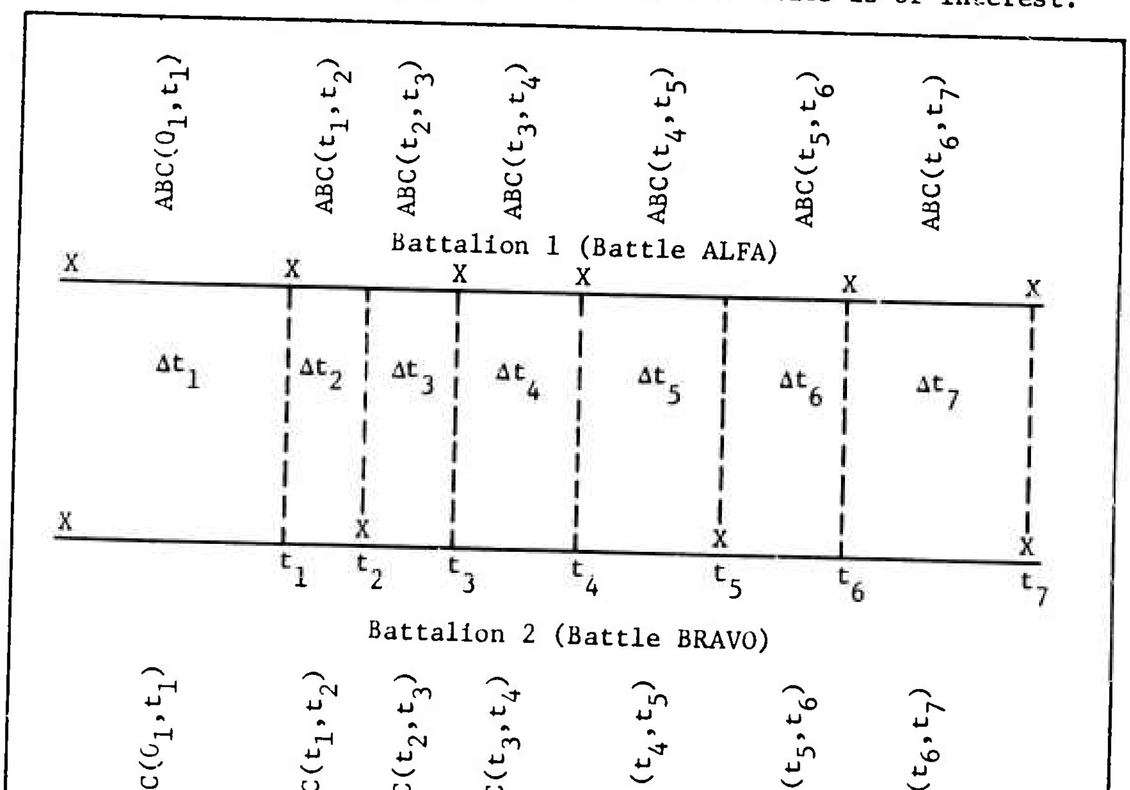


Figure 3-20. Time Dependent Casualty Generation for Battles ALFA and BRAVO

d. The number of casualties for the first brigade in each interval can be extracted from the data on file. For example, during the interval $t=0$ to $t=t_1$ the total number of Brigade 1 casualties is $C_1 = ABC(0_1, t_1) + BBC(0_1, t_1)$. For the increment Δt_2 the number C_2 has a value equal to the number of Blue casualties that occur in Battle ALFA from t_1 to t_2 plus the number in Battle BRAVO from t_1 to t_2; i.e., $C_2 = ABC(t_1, t_2) + BBC(t_1, t_2)$.

e. These numbers can be used to calculate a rate for each time interval Δt. This is accomplished by division and produces seven casualty rates:

$$r_1 = \frac{C_1}{\Delta t_1}$$

$$r_2 = \frac{C_2}{\Delta t_2}$$

$$r_3 = \frac{C_3}{\Delta t_3}$$

$$\vdots$$

$$r_7 = \frac{C_7}{\Delta t_7}$$

Similarly, rates for battles CHARLIE and DELTA (Brigade 2) can be acquired.

f. The next step in the analysis lies in the comparison of rates for two brigades. The logical flow for this comparison is illustrated by Figure 3-21. Stepwise, the procedure used is to:

(1) Acquire game output data for the brigades to be compared.

(2) Select a brigade and battalion.

(3) Order all points in time where a significant event occurs.

(4) Select an effectiveness indicator.

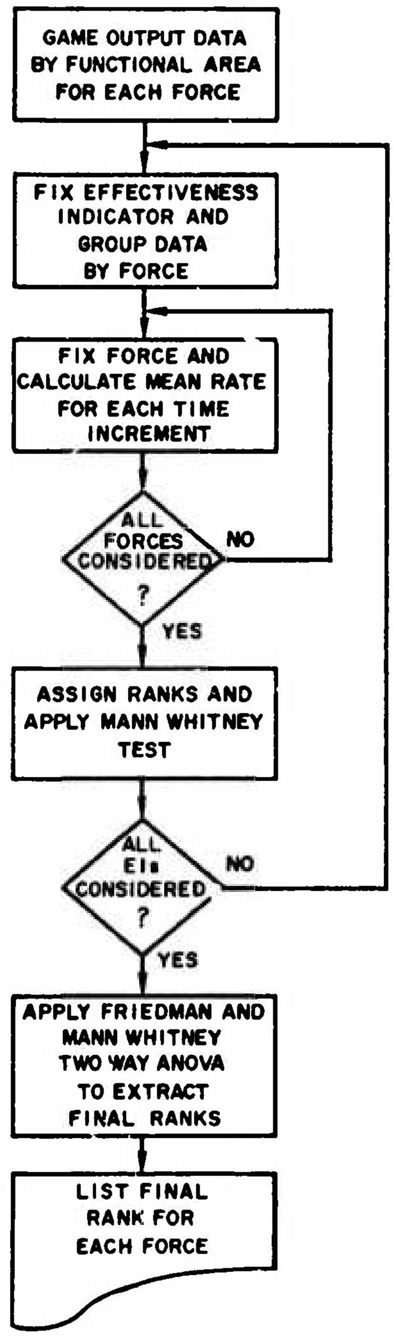


Figure 3-21. Logic Flow for Aggregated Unit Comparisons

(5) Calculate the value of this indicator in each time increment.

(6) Draw a line; i.e., create a computerized file of ordered data, noting each event and the numerical mean value of the effectiveness indicator.

(7) Is this the last battalion of interest?

(a) If not, change to another battalion, note the time dependence of the effectiveness indicator and go to (5).

(b) If yes, continue.

(8) Draw each line (or create a data file) with rates and times noted for each battalion.

(9) Combine battalion rates into a single brigade rate.

(10) Iterate Steps 2 through 9 until all brigade rates have been extracted.

(11) Record rates for Brigades 1 and 2 (Figure 3-22).

Brigade 1	R_1	R_2	R_3	R_4	R_5	R_6	R_7
Brigade 2	r_1	r_2	r_3	r_4	r_5	r_6	r_7

Figure 3-22. Rates for Brigades 1 and 2

(12) Assign ranks to rate data.

(13) Apply the Mann-Whitney test to determine the most effective brigade.

(14) Note the final brigade rank.

(15) If last effectiveness indicator, go to (17).

(16) Change effectiveness indicator and go to (4).

(17) Array the ranks for each brigade by effectiveness indicator.

(18) Apply Friedman two-way ANOVA to test hypothesis H_0: No difference in brigades for this force.

(19) If the null hypothesis must be rejected apply the Mann-Whitney two-way test to determine the most effective brigade.

g. The result of this comparison will be a set of ranks by functional area for the brigades of each force as illustrated in Figure 3-23. With these ranks a subjective analysis of force dependent brigade performance can be initiated.

Force \ Function	Firepower	Mobility	Intel & Control	Combat Service Support
Brigade 1	1	2	1	1
Brigade 2	2	1	2	2

Force I

Force \ Function	Firepower	Mobility	Intel & Control	Combat Service Support
Brigade 1	2	1	1	2
Brigade 2	1	2	2	1

Force II

Figure 3-23. Ranks for Constituent Brigades of Each Force Across Functional Areas

17. BRIGADE COMPARISON ACROSS FORCES:

a. To be consistent with past comparative methodology treatments it is now desirable that the analysis of brigade performance be carried out across forces. The procedure for this comparison is, except for two steps, exactly the same as that presented in paragraph 16f. The difference occurs at Step 11. At this point the analysis of brigades produces an array similar to that shown in Figure 3-24.

B_1	R_1	R_2	R_3	R_4	R_5	R_6	R_7	Force I
B_2	r_1	r_2	r_3	r_4	r_5	r_6	r_7	
B_1	R_1'	R_2'	R_3'	R_4'	R_5'	R_6'	R_7'	Force II
B_2	r_1'	r_2'	r_3'	r_4'	r_5'	r_6'	r_7'	

Figure 3-24. Brigade Rates

The new or different steps are:

(1) Step 11a. Record rates for brigades across Forces I and II.

(2) Step 13a. Apply the Kruskal-Wallis and Mann-Whitney one-way ANOVA to the data presented.

b. The results of this analysis will be a set of brigade ranks across forces. If B_{ij} is used to represent Brigade i in Force j, and it is assumed that only two brigades are important for each force, the methodology will produce a typical set of ranks as illustrated in Figure 3-25 for two forces.

Function / Force	Firepower	Mobility	Intel & Control	Combat Svc Spt
B_{11}	2	1	2	1
B_{12}	1	2	1	3
B_{21}	4	3	3	2
B_{22}	3	4	4	4

Figure 3-25. Brigade Ranks by Functional Area Across Forces

18. TOTAL FORCE COMPARISONS:

a. The methodology for the comparison of two forces differs from that presented for the brigade level comparison within a single force in that it is necessary that total force data and rates be used rather than brigade data. Total force data are acquired by summing the contributions of subordinate battalions. Thus, it is necessary that a file for the total force be created from the battle (battalion level) files. Proceeding as before, there are now four files (or lines) that must be aggregated. For Blue personnel casualties there will be a situation similar to that depicted in Figure 3-26. Total casualties during the increment Δt_1 are $C_1 = ABC(0_1,t_1) + BBC(0_1,t_1) + CBC(0_1,t_1) + DBC(0_1,t_1)$. The rate for this interval is $C_1/\Delta t_1$ as before. Every time interval has a unique rate associated with it, and the set of rates now pertain to the entire force.

b. The comparison of two forces by functional area proceeds as did the comparison of two brigades within a single force. The verbal flow presented in paragraph 16f must be modified only slightly to accommodate this alteration. The logical flow remains unaltered. Verbally, the flow is as follows (assuming only two force are of interest):

(1) Acquire game data for the forces to be compared and select a functional area of interest.

(2) From battalion data files note significant times.

(3) Order all times from smallest to largest: $t_1 < t_2 < t_3 < t_4 \cdots < t_n$.

(4) Calculate the first rate using $ABC(0_1,t_1) + BBC(0_1,t_1) + CBC(0_1,t_1) + DBC(0_1,t_1) \div \Delta t_1 = r_1$.

(5) Note in which battle t_2 occurs (assume battle CHARLIE).

(6) Determine the number of casualties in each battle from time t_1 to time t_2,

(7) Calculate the rate of casualty generation using $ABC(t_1,t_2) + BBC(t_1,t_2) + CBC(t_1,t_2) + DBC(t_1,t_2) \div \Delta t = r_2$.

(8) Iterate through all time intervals.

(9) Record rates.

(10) Change to Force II and repeat Steps 2 through 9.

(11) Assign ranks to rate data.

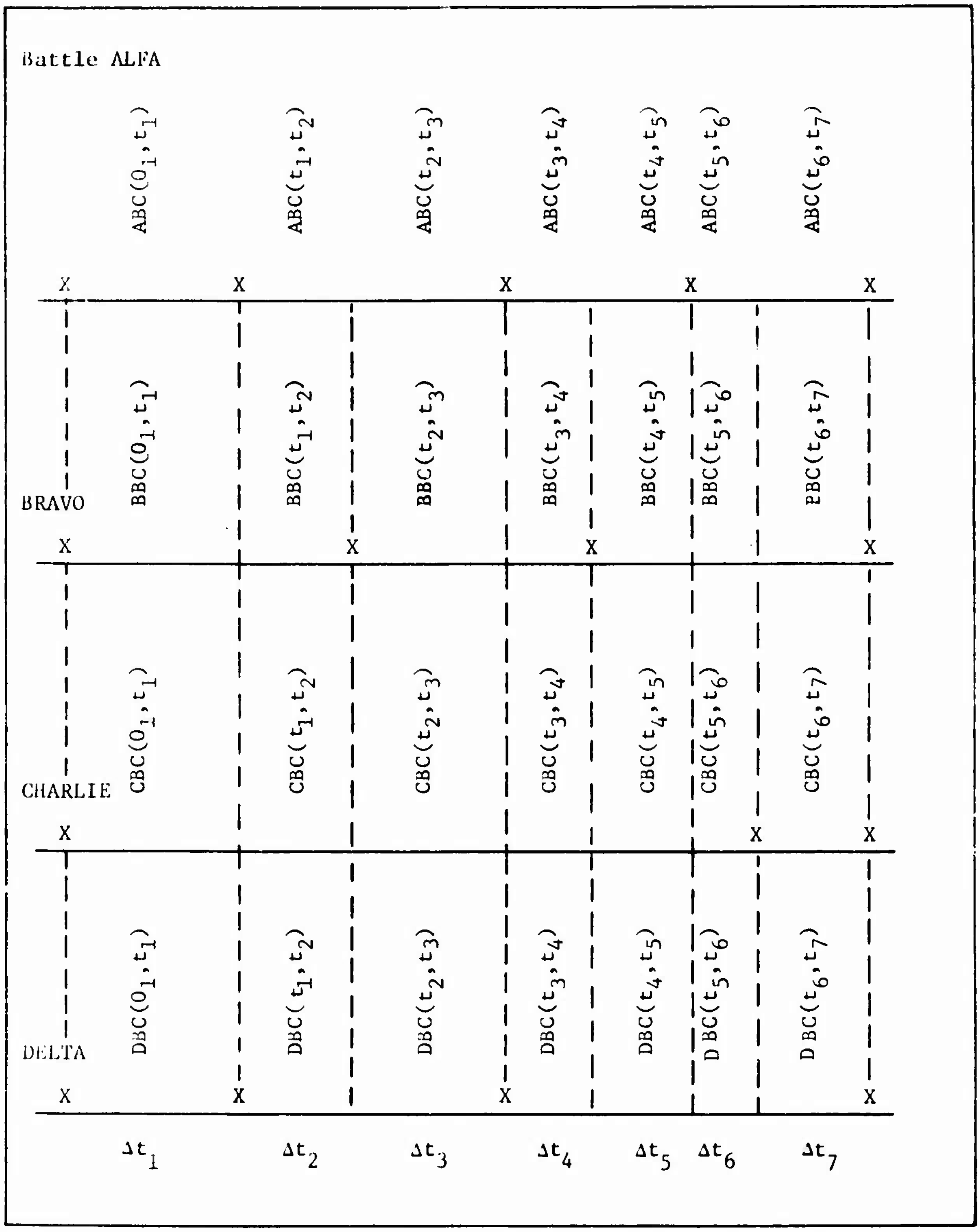


Figure 3-26. Time Dependent Casualty Generation for Four Battles

(12) Apply Mann-Whitney test to determine the most effective force.

(13) Note the rank for each force for this effectiveness indicator.

(14) Repeat Steps 4 through 13 for all effectiveness indicators in this functional area.

(15) Array ranks by effectiveness indicator for this functional area.

(16) Apply Freidman two-way ANOVA to test hypothesis H_0: No difference in forces.

(17) If necessary to reject H_0, apply the Mann-Whitney two-way test to determine the most effective force.

(18) Iterate through all functional areas.

c. One should note that when subordinate unit data are aggregated to total force data, and the analysis applied, there will be produced a set of ranks as illustrated in Figure 3-27.

Function / Force	Firepower	Mobility	Intelligence and Control	Combat Service Support
Force I	1	1	1	2
Force II	2	2	2	1

Figure 3-27. Total Force Ranks by Functional Area

d. To summarize the types of analyses which will be carried out, the approach is one of aggregating subordinate unit data to acquire a set of total force and time dependent data that may be analyzed after combination with like sets from other forces. The analysis of aggregated data allows further input to the overall summary analysis presented in the next section. Also, since this type of analysis tends to relegate a particularly outstanding performance of a single subordinate unit to a portion of the data considered, rather than considering that unit as a member of a certain class of units, the interaction of the subordinate with a weak threat (which produced its outstanding performance data) will be "masked." Such an occurrence is desirable because the treatment of total force data across forces must indicate when total force superiority is really due to the performance of a superior subordinate unit of a given type.

Section V. SUMMARY ANALYSIS

19. INTRODUCTION. This section presents the techniques by which all available subjective and statistical analytical data are assembled, synthesized, and analyzed to compare the combat effectiveness of alternative division forces. It should be recognized that throughout the application of the evaluation and comparison methodologies, a succession of subjective and statistical techniques and schemes have been applied to gain qualitative and quantitative insights into the effectiveness of division forces and their major constituent elements. The principal technique of the summary analysis, then, is to relate and compare analytical results and to make considered judgments as to the relative merits of alternative forces.

20. SUMMARIZATION OF COMPARISONS. As a complement to techniques used in preceding sections of this chapter, quantitative and qualitative performance data are assembled for final subjective analysis. These data relate to unit and system types, to combat activities in which forces engage, and to functions of land combat associated with both unit/system types and combat activities. The subjective analysis is conducted in the following sequence:

a. Primary Measure of Effectiveness. Each division force that has been subjected to gaming and analysis is first judged upon the degree of successful execution of assigned mission. Where alternative forces have all been successful in mission accomplishment, key indicators of the degree of success are arrayed to provide a basis for force comparison. Examples of key indicators are: time to execute mission as a function of time required; the gain (loss) of key terrain as a function of the mission; and total number and hourly rate of Red (Blue) personnel casualties, tank losses, and attack helicopter losses in those cases where the Blue (Red) mission specified attrition or destruction of enemy forces.

(1) Alternative forces that have failed to accomplish their assigned missions are examined in a manner similar to that used for mission-successful forces; however, the principal thrust of the analysis is to identify reasons for nonaccomplishment as well as to analyze the comparative degree of accomplishment (i.e., identify the force that was most nearly successful in accomplishing its mission and determine why).

(2) A force that has been mission-successful may not necessarily be the most combat-effective force overall, and special attention must be paid to the analytical techniques employed to ensure that the combat effectiveness of each force was not unduly influenced by one or a small number of dominant or unbalanced parameters/factors.

(3) The capability of the forces to continue to function in an assigned mission is demonstrated by the residual status of the force. Residual status of units is indicated by personnel strength as a percent of authorized and key equipment item strength as a percent of authorized. The utilization of the residual status indicates whether the force would be capable of further

activities of the type, frequency, and intensity beyond that point in time when the game ceased.

b. Secondary Measures of Effectiveness:

(1) The next step in the summary analysis is to isolate the subjective information and statistical data by secondary measure of effectiveness (land combat functional area). Within each functional area, similar and significant data relating to each division force are arrayed in order that judgments and conclusions relating to functional area performance may be reached. Conflicting data are recognized, analyzed, and resolved. Where more finite comparisons are necessary, graphical presentations and aggregated key information (e.g., killer-victim matrices) will be displayed to support specific judgments. The specific contribution of units and systems to total force effectiveness will be clearly indicated. In some comparisons, it may become necessary to develop a fine structure to the functional area and thereby re-examine effectiveness indicators as producing insights into comparative force effectiveness.

(2) Unusual events and data sets that were not amendable to the mathematical comparison methodology due to a lack of equivalent data for both forces will be presented and submitted to thorough subjective analysis to determine the degree of impact carried by that portion of the data. Similarly, when trends appear at one level of aggregation but are not specifically notable at other levels, this portion of the analysis must point out those differences and analyze reasons for the lack of conformity. Any lack of equivalent trends must be explained by arguments pertaining to the size of data sets available, the sensitivity of the mathematical model that simulates the area of combat, and any force-peculiar gamer technique by which the model for the area of interest was utilized.

c. Side Analyses. The results of any side analyses conducted separately from the main flow of the game will be synthesized. The scope and magnitude of a side analyses can vary; for example, if the dynamic play of the game did not simulate the functions of command, control, and communication (C^3), a study could be made to determine the adequacy of C^3 within the force being gamed under the various situations developed in dynamic play. The results of the side analyses should be subjectively reviewed, integrated in the appropriate analyses, and presented to assist the decision maker.

d. Significant External Factors. Factors external to the actual flow of analysis will be presented, and their influence on combat effectiveness will be analyzed and documented. These factors may include, but are not limited to:

(1) Threat. Differences in threat on the subordinate unit level may cause an outstanding performance or an extremely poor performance of a type unit. When this is the case, the data for opposing forces will be arrayed according to the threat faced and the statistical analyses used to determine if the difference in performance was due to a threat difference.

Using this type of systematic approach, the force performance data will be analyzed across background conditions, and inferences will be tested with regard to the effect on game output of every condition modeled.

(2) Environmental Conditions. The game results, when compared, can lead to large discrepancies between similar forces due primarily to widely varying environmental conditions. These discrepancies in force performance will be reflected in the data, the ranks for subordinate units, and the ranks assigned across brigades and across division forces. Such differences in unit, system, and total force performance may be related for the background environmental conditions; i.e., differences in terrain roughness, differences in light level and vegetation, and differences in prevailing weather conditions. All data with contradictory inferences will be synthesized to determine if these types of background conditions were a significant external factor to the results developed in the analyses.

(3) Reliability/Acceptability of Data. The reliability and acceptability of data used in the study must be highlighted since the uncertainty associated with these data may drive the conclusions of the study. Examples of quantitative uncertainties are hit-kill probabilities, equipment availability rates, ammunition expenditure rates, and reliability statements. The results of sensitivity analyses, which determine if the results are sensitive to the values assigned, must be understood as an influencing factor in making decisions.

(4) Intuitive Judgment. Judgment is used throughout an analysis in the same manner as in the making of an estimate of the situation or a staff estimate. It is necessary in any analysis to use educated guesses and intuitive judgment. These educated guesses and judgments should be identified so that their impact on the conclusions and recommendations are visible.

e. Conclusions and Recommendations:

(1) The conclusions of the study will consider all the pertinent material and data, and the constraints of the study will be recognized and considered in preparing the conclusions. The types of constraints, such as inadequate data base, criticality of assumptions, criticality of uncertainties, and validity of the model, will be clearly pointed out. The conclusions will specifically address the force evaluation objective for which the study was conducted.

(2) Recommendations will be derived logically from the data contained in the study. Recommendations for future war games, analyses, and studies that would facilitiate force comparison will be provided.